

Improving Winter Storm Forecasts with Observing System Simulation Experiments (OSSEs). Part I: An Idealized Case Study of Three U.S. Storms

TANYA R. PEEVEY AND JASON M. ENGLISH

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/OAR/Earth System Research Laboratory/Global Systems Division, Boulder, Colorado

LIDIA CUCURULL

NOAA/OAR/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida

HONGLI WANG^a

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, and NOAA/OAR/Earth System Research Laboratory/Global Systems Division, Boulder, Colorado

ANDREW C. KREN

Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/OAR/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida

(Manuscript received 5 June 2017, in final form 24 February 2018)

ABSTRACT

Severe weather events can have a significant impact on local communities because of the loss of life and property. Forecast busts associated with high-impact weather events have been attributed to initial condition errors over data-sparse regions, such as the Pacific Ocean. Numerous flight campaigns have found that targeted observations over these areas can improve forecasts. To better understand the impacts of measurement type and sampling domains on forecast performance, observing system simulation experiments are performed using the National Centers for Environmental Prediction Global Forecast System (GFS) with hybrid 3D-EnVar data assimilation and the ECMWF T511 nature run. First, three types of simulated perfect dropsonde observations (temperature, specific humidity, and wind) are assimilated into the GFS over a large idealized sampling domain covering the Pacific Ocean. For the three winter storms studied, forecast error was found to be significantly reduced with all three types of measurements providing the most benefit ($\sim 5\%$ – 15% reduction in error). Instances when forecasts are not improved are investigated and concluded to be due to challenging meteorological structures, such as cutoff lows and interactions with atmospheric structures outside the sampling domain. Second, simulated dropsondes are assimilated over sensitive areas and flight tracks established using the ensemble transform sensitivity (ETS) technique. For all three winter storms, forecast error is reduced up to 5%, which is less than that found using an idealized domain. These results suggest that targeted observations over the Pacific Ocean may provide a small improvement to winter storm forecasts over the United States.

1. Introduction

Extratropical cyclones that originate over the Pacific Ocean can impact the West Coast of the United States

with high winds and heavy precipitation and can result in the loss of life or property. Even with advances in numerical weather prediction (NWP), large errors in short-term forecasts of strong North Pacific cyclones still occur (McMurdie and Mass 2004; McMurdie and Casola 2009; Rodwell et al. 2013). For example, McMurdie and Mass (2004) observed large short-term forecast errors in position and central pressure of midlatitude cyclones along the west coast of North America and offshore. Subsequent studies

^a Current affiliation: Panasonic Weather Solutions, Morrisville, North Carolina.

Corresponding author: Tanya Peevey, tanya.peevey@noaa.gov

have shown that short-term forecast errors are larger along the west coast of North America than the east coast (Wedam et al. 2009), and central pressure displacement errors for cyclones are large (greater than ~ 10 hPa on day 2) over the central and eastern Pacific for the Global Forecast System (GFS; Charles and Colle 2009; Colle and Charles 2011).

Errors in short-term forecasts have been partially attributed to initial condition errors that occur over the data-sparse region of the Pacific Ocean (Langland et al. 1999b, 2002) or are tied to flow-dependent model error growth (McMurdie and Mass 2004; Kleist and Morgan 2005). These initial condition errors have been shown to not only affect local forecast accuracy, but also to spread eastward from the Pacific Ocean and impact the eastern United States (see McMurdie and Casola 2009; Colle and Charles 2011, and references therein). Some of the current understanding of initial condition errors was obtained through field campaigns, such as the Winter Storm Reconnaissance (WSR) that started in 1999 and ran for 10+ years. Research results from WSR campaigns have shown that forecasts improve due to data enhancing the analysis over the eastern Pacific Ocean (Toth et al. 2000; Szunyogh et al. 2000). Other studies have shown that certain flow regimes over the United States can influence predictability and forecasts in Europe and suggest that it is important to improve the quality and use of observational data in regions that are ill-constrained (Rodwell et al. 2013).

Part of the effort to reduce the impact of model initial condition errors on the forecast performance of high-impact weather events has occurred through a variety of field campaigns using the concept of “targeted observations.” The goal of a targeted observation effort is to reduce the analysis errors in the model by targeting regions that have the potential for large error growth, thus curtailing the growth of these initial errors and reducing the subsequent forecast error (Toth et al. 2002; Majumdar 2016). The various methods used for finding these target regions are called adaptive observation strategies, such as the ensemble transform (ET; Bishop and Toth 1999; Szunyogh et al. 1999), ensemble transform sensitivity (ETS; Zhang et al. 2016), and adjoint-based techniques (Palmer et al. 1998; Langland et al. 1999a). One study comparing targeting techniques found that forecasts improved regardless of the technique utilized (Aberson et al. 2011), while another study found differences in identified target areas when the forecast metrics were similar (Huang and Meng 2014). These results suggest that targeting techniques are sometimes unreliable. Additionally, other studies have shown that targeted observation methods are more effective at increasing forecast skill than random observations (Morss et al. 2001; Buizza et al. 2007) and result in overall

positive results, even though there are null cases or degraded forecasts (Szunyogh et al. 1999).

Both observing system simulation experiments (OSSEs) and observing system experiments (OSEs) have been used with targeted observation techniques to evaluate the impact of observations on forecast skill. Early OSE data impact studies focused on extratropical storms have shown, using a variety of targeting techniques, that the assimilation of targeted observations reduces forecast error on average for short-term (0–60 h) forecasts (Langland et al. 1999a,b; Szunyogh et al. 2000). A more recent study found neutral to negative results when investigating the data impact of the 2011 Winter Storm Reconnaissance program (WSR2011) on forecast skill in the European Centre for Medium-Range Weather Forecasts (ECMWF) four-dimensional variational (4D-Var) data assimilation (DA) system (Hamill et al. 2013). The authors suggested that results were due to improvements in the forecast and DA system and an increase in the number of observations over the past decade. However, another study found results to be sensitive to the meteorological situation and season, and it was suggested that a 4D-Var system would better exploit localized targeted observations and thus result in greater benefits (Cardinali et al. 2007). Recent reviews of the impact of targeted observations on the forecast skill of polar, extratropical, and tropical cyclones have found the following: neutral results for polar cyclones (Irvine et al. 2009, 2011), small but positive average impact for short-term forecasts of extratropical and tropical cyclones [see Majumdar (2016) and references therein], and a 10%–15% improvement in hurricane track forecasts at the 0–5-day lead time (Aberson 2010) with occasional degradations (Aberson 2008).

OSSEs were first suggested by Newton (1954) as a method for evaluating the impact of future observing systems on numerical forecasts using fictitious observations. They are a cost-effective method to evaluate the potential value of an observing system before it is deployed. Over the last few decades, OSSEs have become increasingly realistic and are performed with operational models (Boukabara et al. 2016). A recent review of the OSSE methodology and its current state can be found in Hoffman and Atlas (2016). OSSEs have been used to evaluate the impact of many types of observing systems on forecast skill and the DA system. Both Atlas et al. (2015) and Ma et al. (2015) evaluated the impact of simulated wind lidar observations on forecast skill and found an improvement in the wind and mass fields, both in the extratropics and tropics. Moreover, Atlas et al. (2015) found that the short-term forecasts over the tropics were significantly reduced. OSSE studies have

also evaluated the impact of rawinsondes globally for wind and temperature fields (Privé et al. 2014a) and dropsondes for tropical cyclones (Privé et al. 2014b). Both studies found an improvement in the short-term forecast skill for wind and temperature fields and tropical cyclone tracks, respectively. Other OSSE studies have been used to evaluate the impact of different DA systems on forecast skill (Kleist and Ide 2015a,b).

Previous studies have found that supplemental observations improve forecast skill, but none have evaluated extratropical cyclones using the targeted observation approach with a modern NWP system in the OSSE framework. This study will be the first to look at the impact of targeted observations, along with observation type, on the forecast skill of Pacific cyclones in the OSSE framework. Additionally, the hypothesis of whether or not targeted observations can improve Pacific cyclone forecasts is tested in this study, and results can provide guidance to mission scientists when planning real field campaigns, such as dropsonde placement.

This study is the first of a two-part study, where Part II looks at the ability of targeted dropsonde observations to mitigate the potential loss of the *Suomi National Polar-Orbiting Partnership (NPP)* satellite and the expected increase in forecast error of high-impact weather events until data from the replacement satellite, *NOAA-20* (launched in November 2017), is used by global models (English et al. 2018). Here, we use OSSEs to quantify the impacts of targeted dropsonde observations, deployed over the Pacific Ocean, on the forecast accuracy of three high-impact weather events over the United States. A case study makes establishing statistical significance difficult since results can vary with storm, forecast hour, initialization date, and verification region. This highlights why there is still some debate as to the benefit of targeted observations and the need for more research in this area. There are three primary objectives of this study: 1) investigate the relative importance of different types of dropsonde measurements (temperature, wind, and specific humidity) on forecast error, 2) investigate the impact of different Pacific Ocean sampling domains on forecast error, and 3) conduct an in-depth analysis of select cases to understand the meteorological features impacting results with targeted observations.

2. Data and methodology

a. OSSE framework and setup

An OSSE system contains the following components: 1) a long atmospheric model integration using a “state of the art” system considered to be the “truth” or nature run (NR) that represents the real atmosphere (e.g.,

possesses a realistic climatology, realistic storm track patterns); 2) observations simulated from the NR using the same distribution and accuracy characteristics as real-world observations; 3) simulation of new observations with realistic accuracy and coverage; 4) a different DA/forecast system, which is typically set to a lower resolution than the NR, to ingest the simulated observations; and 5) validation of the entire OSSE system (Hoffman and Atlas 2016). OSSEs can be used to evaluate the impact of global, regional, or mesoscale observations on NWP.

The NR used in this study, T511 NR, is a 13-month free-running forecast with T511 horizontal resolution (about 40 km) and 91 model levels and was produced in 2005 by ECMWF using their Integrated Forecast System (IFS), version cy31r1. The forecast was initialized at 1200 UTC 1 May 2005, with the operational analysis as the initial conditions, and ran until 0000 UTC 1 June 2006, with output every 3 h (Masutani et al. 2007; Andersson and Masutani 2010). The listed year is somewhat arbitrary since this is a simulated atmosphere. The only link to the real atmosphere for 2005–06, aside from initial conditions, is through the lower boundary conditions (sea surface temperatures and sea ice data) (Riishøjgaard et al. 2012). For further details on the ECMWF model, see Errico et al. (2013) and references therein. The T511 NR was validated, and it was found that hurricanes, midlatitude cyclones, midlatitude jets, and cloud statistics were realistic (Masutani et al. 2007; Reale et al. 2007; Andersson and Masutani 2010; McCarty et al. 2012; Errico et al. 2013).

In an OSSE, appropriate and realistic errors should be added to simulated observations to properly simulate real-world observation statistics. For a detailed explanation of the various sources of error in the context of OSSEs, see Errico et al. (2013). In this study, perfect simulated observations are used and, therefore, are a limitation of the study (Masutani et al. 2010). However, the goal of this study is to evaluate the impact of observations in data-sparse regions, and using simulated perfect observations satisfies this goal. Moreover, there is an advantage to using perfect observations; any impact on the forecast skill could be greater than what would be seen in the real world due to the observations. It should be noted that the perfect simulated observations will be prescribed uncertainties when assimilated by the Gridpoint Statistical Interpolation analysis system (GSI). More information on the magnitudes of these uncertainties can be found in the GSI user's guide (Hu et al. 2017).

Perfect simulated observations are generated from the NR for January and February 2006 at times and locations of actual observations during 2012. This approach leads to observations with realistic spatial distributions that do not consider the NR cloud field.

TABLE 1. List and description of the different experiments performed in this OSSE study.

Expt	Name	Dropsonde obs	Description
Control	CTL	None	Control experiment, Jan–Feb 2006, includes simulated conventional and satellite observations (see Table A1)
Idealized	Ideal_tquv	Temperature, specific humidity, and wind	CTL plus simulated dropsonde observations over a large domain over the northern Pacific Ocean
	Ideal_t	Temperature	
	Ideal_q	Specific humidity	
	Ideal_uv	Wind	
Sensitivity	Sensit_tquv	Temperature, specific humidity, and wind	CTL plus simulated dropsonde observations over a smaller domain over the northern Pacific Ocean shown as regions sensitive to error growth determined by the ETS method
Flight	Flight_tquv	Temperature, specific humidity, and wind	CTL plus simulated dropsonde observations over a flight path that is informed by the smaller domain established in the Sensitivity experiment

This can result in simulated observations in unrealistic locations ([Boukabara et al. 2016](#)), and so the distribution of the NR cloud field in the region of simulated dropsonde observations was examined. It was found that the percentage of cloudy to cloud-free grid points was similar for all experiments for the two case studies of primary focus in this manuscript. Therefore, issues with cloud-affected observations should not significantly alter the presented results.

Simulated satellite observations were generated using the Community Radiative Transfer Model (CRTM, version 1.2.2) and temperature and specific humidity observations from the NR. A newer version of the CRTM, version 2.2.1, is used by the assimilation system employed in this study. Conventional observations (wind, temperature, surface pressure, and specific humidity) were simulated by sampling the T511 NR at the location and time given by the National Centers for Environmental Prediction (NCEP) for that observation. For more details on the generation of simulated satellite and conventional observations used in this study, see [Zhu et al. \(2012\)](#), and for access to the data, visit the Joint OSSE website (<http://jointosse.org/>). For reference, the dataset used in this study was also used in [Riishojgaard et al. \(2012\)](#) and [Ma et al. \(2015\)](#). A list of satellite and conventional observations assimilated by the NCEP Global Data Assimilation System (GDAS) for this study can be found in [appendix A](#) in [Tables A1](#) and [A2](#), respectively. The list of satellites was motivated by the need to be as close to operations as possible so the conclusions would be more useful to society. As a result, the exact satellite instruments assimilated were determined by the version of the operational NCEP system selected for this study.

For all three experiments (Idealized, Sensitivity, and Flight) described in [Table 1](#), dropsonde observations of

wind, temperature, and specific humidity are simulated from the T511 NR. They do not drift with the wind and, therefore, should be considered model soundings. Dropsondes are released in the sampling domain every $\pm 1^\circ$ of latitude–longitude from 70 hPa and descend to the surface crossing the following pressure levels: 70, 100, 150, 200, . . . , 650, 700, 725, 750, . . . , 975, and 1000 hPa. These pressure levels go beyond, but still include, mandatory levels that are greater than and equal to 70 hPa. Lower pressure levels are not included due to the experiment design that requires the simulated dropsondes to mimic an actual flight campaign with the Global Hawk (GH), which has an altitude limit of 19.8 km. Flight path determination is discussed in [section 3](#).

In this study, the analysis/forecast system used to assimilate the observations is the NCEP GDAS Q1FY15 operational implementation using the 3D variational (3D-Var) configuration with the hybrid ensemble Kalman filter (EnKF) ([Kleist and Ide 2015a](#)). Specifics of the FY15 implementation can be found using the URL in the [McClung \(2014\)](#) reference. Both the GSI and GFS components are configured with T382 horizontal resolution and 64 vertical levels (about 50 km). The EnKF component is set up with T254 horizontal resolution and 64 vertical levels. The GSI hybrid ensemble–variational DA system is used to initialize the GFS model. The theoretical framework, implementations, and tests of the hybrid system can be found in [Wang \(2010\)](#), [Kleist \(2012\)](#), [Wang et al. \(2013\)](#), and [Kleist and Ide \(2015a\)](#). The flow-dependent background error covariance is estimated using a serial square root filter form of an EnKF developed by [Whitaker and Hamill \(2002\)](#) and [Whitaker et al. \(2008\)](#). An ensemble with 80 members is employed in this research. The weights given to the static (fixed) and ensemble-based flow-dependent background error covariances are 0.125 and

0.875, respectively. The horizontal localization length scale for the hybrid system and EnKF is level dependent, increasing from 350 km at the surface to 1300 km at the model top. Multiplicative inflation (Whitaker and Hamill 2012) is employed to maintain ensemble spread, with the multiplicative factor set to 0.85. This model configuration is used for all experiments, which are supported by the Sensing Hazards with Operational Unmanned Technology (SHOUT) project. The SHOUT project tests the impact of targeted observations from unmanned aircraft systems (UASs), specifically the GH, on the forecast performance of high-impact weather events from NWP systems (Kren et al. 2018, manuscript submitted to *Wea. Forecasting*).

The OSSE system used in this study was validated by comparing our results to OSE results that used the same model configuration (Kren et al. 2018, manuscript submitted to *Wea. Forecasting*). A comparison of globally calculated anomaly correlation coefficients versus forecast hour showed that the OSSE system behaved in a similar manner to that of the OSE. Additionally, the minimum sea level pressure (SLP) and 6-h accumulated precipitation of the three storms found in the NR (described in section 4a) were compared to what is typical for real-world winter extratropical cyclones and were found to be realistic. For reference, simulated datasets produced from the T511 NR with the same methodology have been used by previous researchers (Riishojgaard et al. 2012; Ma et al. 2015; Atlas et al. 2015; Cucurull et al. 2017) and, in some cases, compared to corresponding OSEs for calibration purposes. In those studies, the spectral resolution of the forecast model was set to T382 in a 2012 version of the NCEP GFS model.

b. Total energy error calculation

The impact of additional observations on forecast skill is quantified using an approximation of the dry total-energy norm, hereafter total energy error (TEE), a well-known formula that provides a comprehensive measure of forecast error by taking into account temperature and wind errors at multiple vertical levels (Hamill et al. 2013). In this study, we calculate energy error (in meters per second) at 200, 500, and 700 hPa, consistent with the vertical levels used by the ETS technique (Zhang et al. 2016). Near-surface fields considered by Hamill et al. (2013) are not used here because their analysis errors are small over the Pacific Ocean, the region of focus for this study, and for the reasons presented in appendix B. Additionally, the dry total-energy norm is used instead of the moist energy norm because moist processes, such as convection, occur at scales smaller than the resolution of the model used here, and moist processes have been shown to result in targeting areas that are less reliable

than those found with TEE (Huang and Meng 2014). Equation (1) shows the formula used in this study:

$$E = \left\{ \frac{1}{2} \int_A \left[\frac{1}{3} \left(u_{200}^2 + v_{200}^2 + \frac{c_p}{T_r} t_{200}^2 \right) + \frac{1}{3} \left(u_{500}^2 + v_{500}^2 + \frac{c_p}{T_r} t_{500}^2 \right) + \frac{1}{3} \left(u_{700}^2 + v_{700}^2 + \frac{c_p}{T_r} t_{700}^2 \right) \right] \right\}^{1/2}, \quad (1)$$

where u , v , and t represent the difference between the forecast and the NR for the zonal wind, meridional wind, and temperature, respectively. Additionally, T_r is the reference temperature (300 K), and c_p is the specific heat of dry air at a constant pressure ($1004 \text{ J K}^{-1} \text{ kg}^{-1}$). The verification region in this study is represented by A and thus, with the integral, calculates the average error over domain A . In section 4, domain A is either the western United States (WEST_USA: $30^\circ\text{--}50^\circ\text{N}$, $235^\circ\text{--}260^\circ\text{E}$) or Alaska (ALASKA: $55^\circ\text{--}70^\circ\text{N}$, $195^\circ\text{--}220^\circ\text{E}$). This verification domain is chosen because 1) the assimilation of observations over the Pacific Ocean (see Fig. 2) should impact forecast error downstream of that location (i.e., the United States), and 2) the impact could extend beyond the storm of interest. In section 5, domain A is smaller and consists of $14^\circ \times 14^\circ$ boxes over the contiguous United States (CONUS). This smaller verification domain is used for the purpose of isolating and analyzing the impact of the observations on the storm itself and not the surrounding environment.

3. Experiment design

A series of experiments was conducted to understand the impacts of dropsonde measurement type (temperature, wind, and specific humidity) and dropsonde sampling domain on forecast error. The list of experiments and their descriptions are presented in Table 1. The control (CTL) was initialized at 1800 UTC 1 January 2006 using NCEP production run files and cycled for 2 months (1 January–28 February 2006). Both simulated satellite and conventional observations listed in Tables A1 and A2 were assimilated during this time, with the first 10 days considered the spinup period. Plots of Northern Hemisphere analysis error (CTL NR) for geopotential height, temperature, and wind at multiple pressure levels were used to find this spinup period. Additionally, all cases studied here occur well beyond this spinup period. For the three experiments described in Table 1 (Idealized, Sensitivity, and Flight), dropsonde observations were assimilated over the northeastern Pacific Ocean, in addition to observations in the CTL,

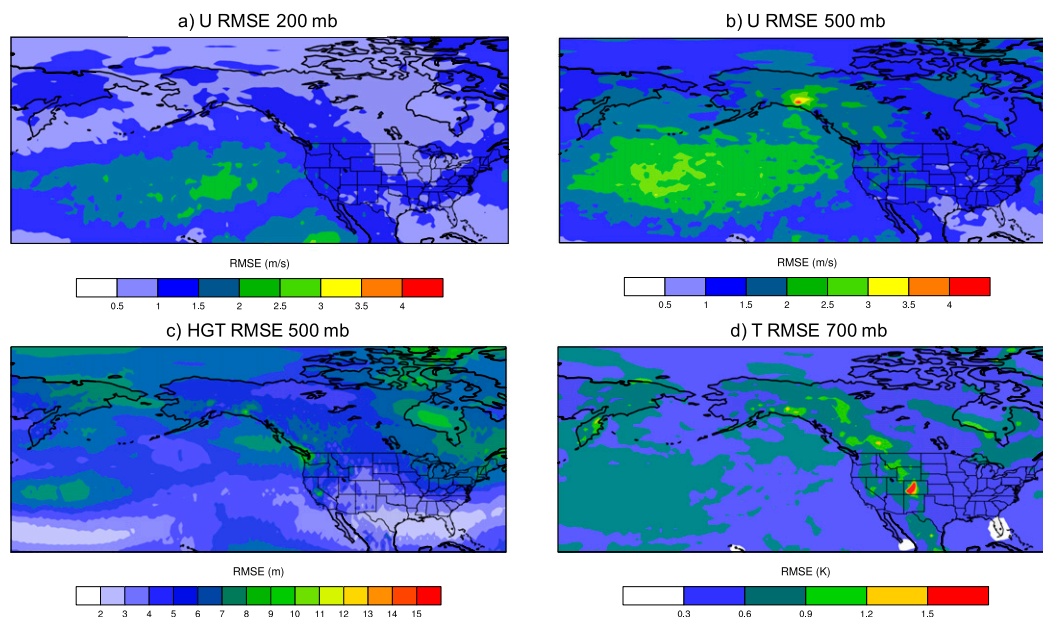


FIG. 1. Map of analysis RMSE (CTL vs T511 NR) over the northeastern Pacific for (a) 200-mb zonal winds, (b) 500-mb zonal winds, (c) 500-mb geopotential heights, and (d) 700-mb temperatures. All fields are calculated and averaged between 10 Jan and 28 Feb 2006 and include every 6-h analysis cycle (i.e., four cycles per day).

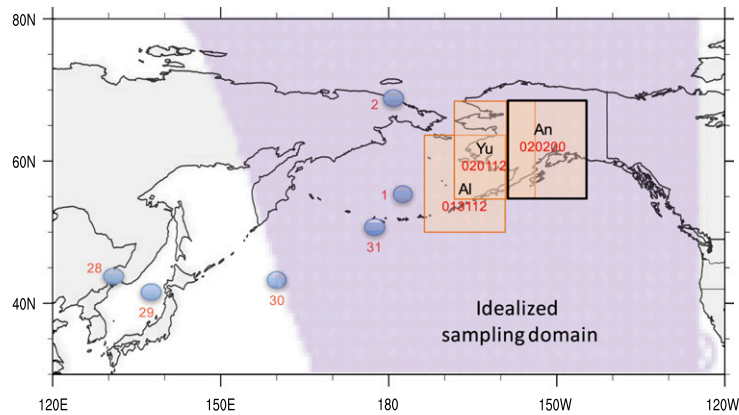
and cycling was initiated using initial conditions from the CTL. The northeastern Pacific Ocean was chosen for multiple reasons: 1) it was an area of interest under SHOUT, 2) large analysis errors are typically found in this region during the winter (see section 1), and 3) significant analysis errors were found in this region when comparing the CTL to the T511 NR (see Fig. 1). Figure 1 includes both fields that are used in the TEE formula and fields that have meteorological significance. Root-mean-square error (RMSE) is used as the metric since it is a well-established method for showing the difference in values between a population and a sample.

For the Idealized experiments, perfect dropsonde observations (temperature, wind, and specific humidity) were simulated at every grid point (data density of $\pm 1^\circ$ of latitude–longitude) within the idealized domain (purple region defined in Fig. 2) and then assimilated into the NCEP GFS system. For the Sensitivity and Flight experiments, maps were generated using the ETS method to identify regions sensitive to large error growth and to aid in the placement of simulated dropsonde observations. In the ETS method, the gradient of the forecast error variance to analysis error variance is calculated for each predefined verification region. As an approximation of the ET method (Bishop and Toth 1999), it has the advantage of only requiring the calculation of a single transform matrix, making it more computationally efficient. More information on the ETS technique and its benefits can be found in Zhang et al. (2016) and

Majumdar (2016). The ETS sensitivity was determined from forecast data that were available approximately 5 days prior to the storm, reaching the predefined verification region at the verification time (see Fig. 2). ETS sensitivity maps were then generated to identify targeting regions for the 2–3-day lead time for each verification region. Dropsonde releases occur within the 2–3-day time window at 0000, 0600, 1200, and 1800 UTC center times of the analysis cycles, for a total of five cycles with the same data density as in the Idealized experiments.

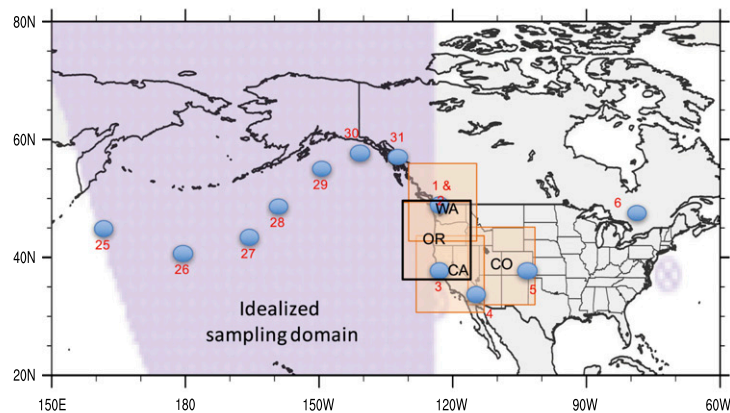
All UAS flight paths are determined using an automated flight track algorithm (appendix C). It was designed to take into account the gradient in the ETS sensitivity maps generated for each case and the GH flight requirements listed in appendix C. The final flight path, which determines dropsonde release locations, starts at the 3-day lead time relative to the verification time at a waypoint determined by the algorithm and continues until the 2-day lead time, with dropsondes released for a total of five cycles. This setup is designed to mimic an actual GH flight campaign with a $\pm 1^\circ$ along-track spacing of dropsonde locations.

Exact dates of cycling for each experiment are discussed in the pertinent sections. GFS forecasts were initiated using the analysis from their own GDAS cycle as their initial conditions. For each experiment, dropsonde observations of temperature, wind, and specific humidity are simulated with varying data densities

a) Track of January 29th Storm

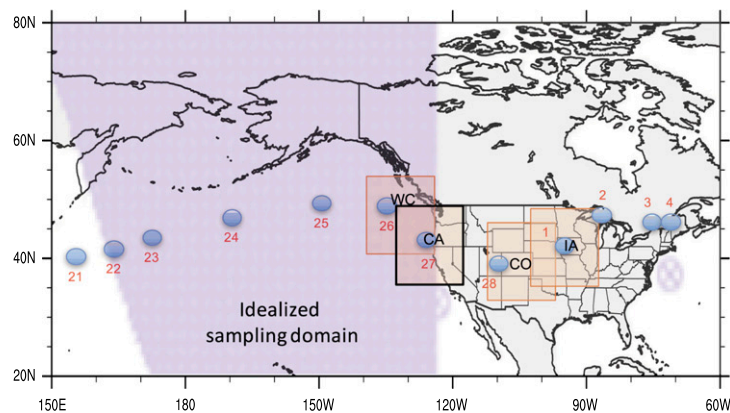
All dates are at 00Z unless noted otherwise.

b) Track of January 30 Storm



All dates are at 00Z unless noted otherwise.

c) Track of February 25 Storm



All dates are at 00Z unless noted otherwise.

FIG. 2. Schematic of storm tracks for the (a) 29 Jan, (b) 30 Jan, and (c) 25 Feb storms. Blue circles represent the approximate center of the storm, and the numbers below show the date when the storm was at that location. All dates are for 0000 UTC unless noted otherwise. The purple region represents the idealized domain, the domain in which perfect dropsondes observations (t , q , and uv) were simulated for the Idealized experiments. The orange boxes represent the four different $14^\circ \times 14^\circ$ lat–lon verification regions used to analyze observation impact on forecast performance, and the boxes with the black outline represent the verification region used to generate the ETS maps.

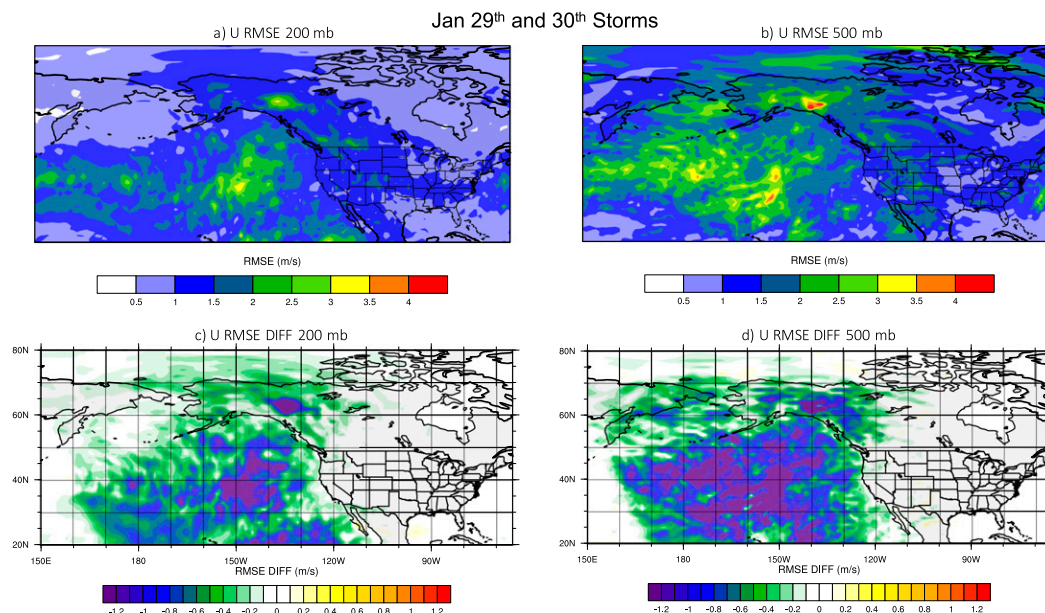


FIG. 3. (top) RMSE in the analysis (w.r.t. T511 NR) and (bottom) change in the analysis RMSE after dropsondes are assimilated (Idealized – CTL) over the northeastern Pacific for (a),(c) 200- and (b),(d) 500-mb zonal wind. Plots include analysis fields for every cycle (four cycles per day) between 0000 UTC 25 Jan and 1200 UTC 2 Feb (29 and 30 Jan storms).

specified by each experiment and placed over a region that is both vertically and horizontally undersampled. These same fields (except specific humidity—reasons discussed in section 2b), which the dropsonde observations are targeted to improve, are in the TEE formula that is used in the ETS technique for targeting and in the evaluation of forecast error.

4. Results: Idealized experiments

a. Storm characteristics

Over the 2-month period of the CTL, three storms were identified in the T511 NR, based on the following criteria: 1) they had developed or passed through the northern Pacific Ocean on their way to the West Coast of the United States; 2) they were long-lived (lasting a minimum of 4 days); 3) they had a well-defined structure in the 500-mb (1 mb = 1 hPa) geopotential height field (low heights and the formation of a closed contour); and 4) they had at least 4–5 mm of precipitation over a 6-h period at any grid point within a $14^\circ \times 14^\circ$ box surrounding the storm when it was over the United States. The identified storms are denoted as the 29 January storm, 30 January storm, and 25 February storm, respectively, with the track of each storm shown in Fig. 2.

The 29 January storm began off the southernmost portion of the east coast of Russia as an upper-level low passed over a baroclinic region. The upper-level low

supported the formation of a surface low that intensified as it moved eastward. Around 30–31 January, the low pressure system merged with another storm originating off the coast of Russia and intensified, reaching a minimum SLP of 956 mb. The system continued to move toward the northeast into the Bering Sea, bringing precipitation to south-central and southwest Alaska. Cycling for this storm for each experiment listed in Table 1 was initiated at 1800 UTC 24 January and terminated at 0000 UTC 5 February, with forecasts initiated every 0000 and 1200 UTC between 0000 UTC 25 January and 1200 UTC 2 February.

The 30 January storm formed ahead of the 29 January storm from a downstream upper-level low. As this upper-level system moved into the Gulf of Alaska, a closed low formed aloft and remained with the downstream ridge as it moved across the southern part of Alaska. The closed low weakened and partially dissipated as it tracked south along the coast of Canada and into the United States. Around 30 January, it began to strengthen again, and it then intensified further around 1 February due to a short wave originating upstream along the ridge that was subsequently amplified by an upper-level low over Canada. The 30 January storm then moved farther south, into the northwestern United States and then along the California coast, bringing small amounts of precipitation to both regions and obtaining a minimum SLP of 998 mb. The storm then

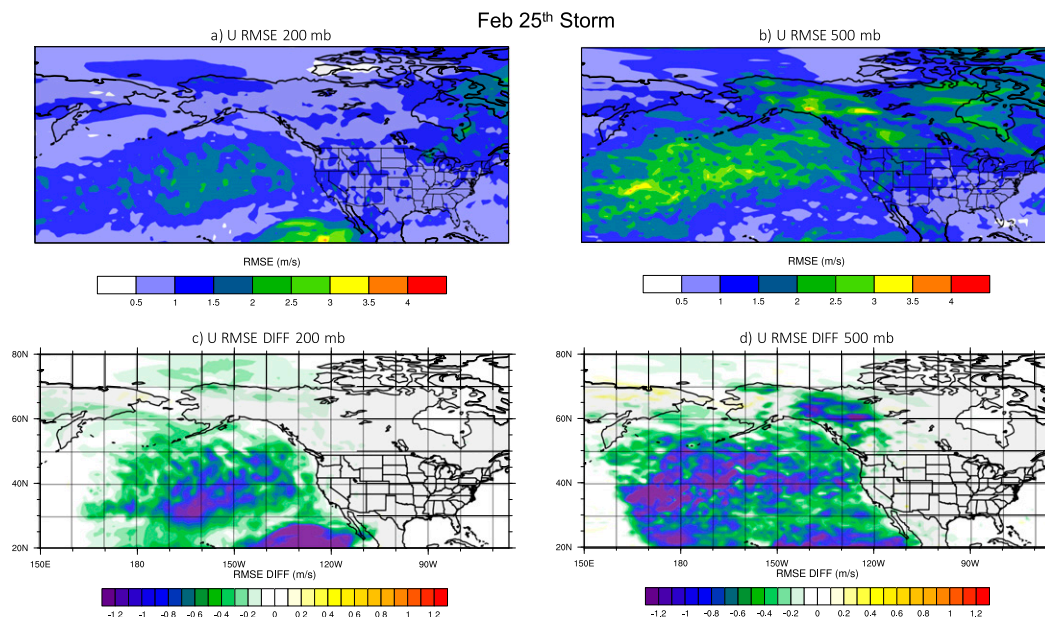


FIG. 4. As in Fig. 3, but between 0000 UTC 20 Feb and 0000 UTC 28 Feb (25 Feb storm).

crossed Arizona and New Mexico and began to merge with the mean flow and dissipate. The cycling and forecasting period that was used for the 29 January storm was also used for this storm.

The 25 February storm originated as a closed low off the coast of Japan, traversing eastward across the Pacific Ocean toward Alaska. The closed low formed aloft on 25 February and reached the northwestern United States around 27 February, embedded within an upper-level ridge. During those 2–3 days, the storm strengthened, reaching a minimum SLP of 983 mb and bringing precipitation to Washington, Oregon, and parts of Northern California. Twelve hours after reaching the United States, the storm exited the coast of California and penetrated inland, rejoining the mean flow. As it moved across Colorado and into the Midwest, the storm began to weaken. A short wave over southeastern Canada then caused it to intensify again and form a cold frontal boundary stretching from Wisconsin to eastern Texas. Cycling for this storm for each experiment listed in Table 1 was initiated at 1800 UTC 19 February and terminated at 1800 UTC 28 February. Forecasts were initiated every 0000 UTC and 1200 UTC between 0000 UTC 20 February and 0000 UTC 28 February.

b. Analysis and forecast performance

Before initiating cycling with added dropsonde observations, the analysis error from the CTL was examined. Figures 3a,b and 4a,b have large analysis RMSE over the central and eastern portions of the northern

Pacific Ocean at the 200- and 500-mb pressure levels, which highlights a region with the potential for a reduction in the analysis error following the assimilation of dropsonde observations. The location of the larger analysis errors is spatially consistent with previous studies (McMurdie and Mass 2004; McMurdie and Casola 2009). Temperature at 700 mb and geopotential height at 500 mb were also examined (not shown) and were found to have the same general pattern of RMSE shown in Fig. 1. After assimilating the dropsonde observations into the NCEP GFS system, the analysis error was examined (Figs. 3c,d and 4c,d) and was found to be reduced in the idealized domain (purple area highlighted in Fig. 2), verifying the impact of the dropsonde observations on the analysis.

Following the multiday cycling for each Idealized experiment, the forecast error is examined using the formulation of the TEE equation described in section 2b. Results for each storm are shown in Fig. 5. In addition to the percent mean difference (solid lines), this figure shows the 95% confidence interval established using the paired t test (dashed lines), where the two populations being compared are each experiment versus the CTL. Therefore, results are statistically significant when the solid line is beyond the bounds established by the corresponding dashed line. Variability is high in Fig. 5 due, in part, to averaging across a broad range of forecast lead times (0–7 days). These figures are produced using data from the spinup period of the Idealized experiments since the conclusions do not change by excluding that period. For Figs. 5 and 6 (and Figs. 12, 13),

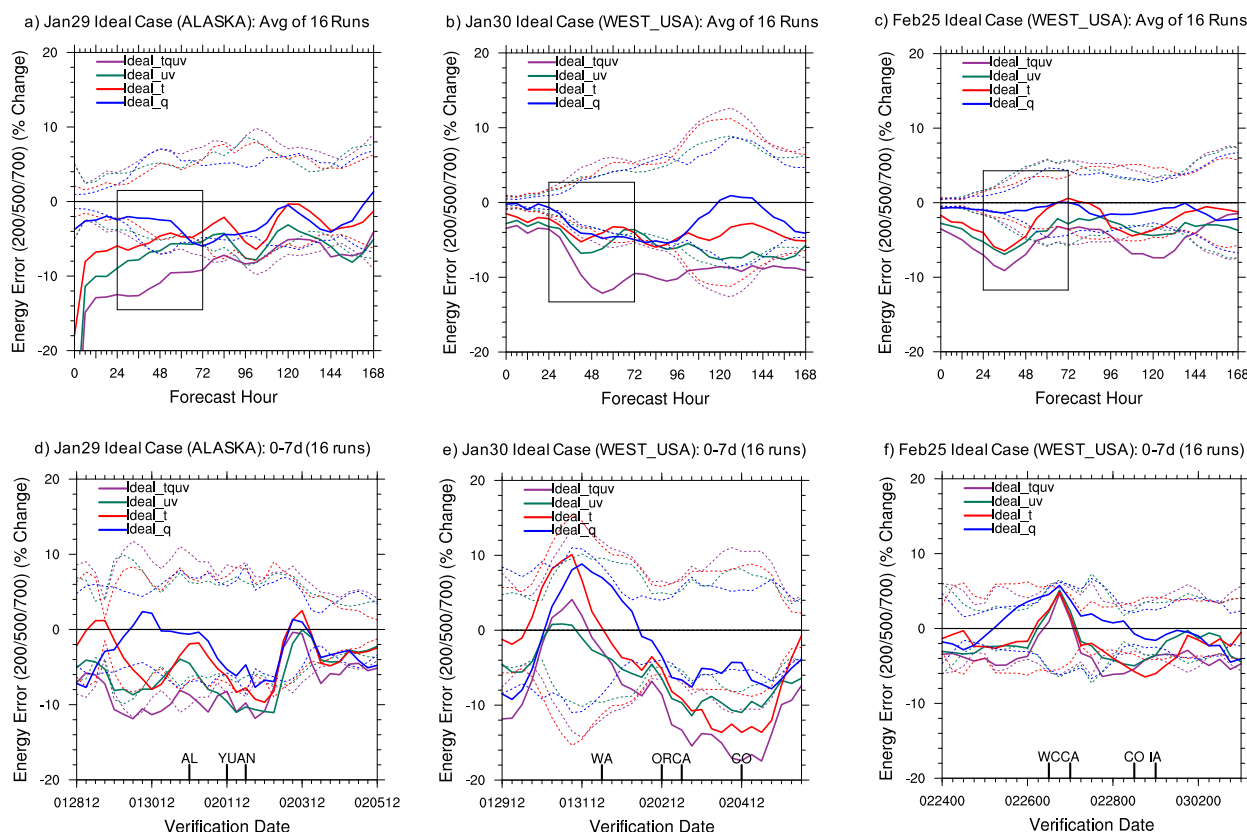


FIG. 5. TEE for each storm (columns) over the western United States for the 25 Feb and 30 Jan storms and over Alaska for the 29 Jan storm. Error values are generated by calculating TEE for all forecasts with a 0–7-day lead time then averaging the error for each storm over the domain of interest (either WEST_USA or ALASKA). This metric is calculated with respect to (top) forecast hour and (bottom) verification date. The solid lines for each case study represent the percent difference between the CTL and the experiments: Ideal_tquv (purple), Ideal_uv (green), Ideal_t (red), and Ideal_q (blue). The corresponding dashed lines represent the 95% confidence interval obtained using the paired t test, with one population being an experiment and the other the CTL. Boxes in the top row highlight the 1–3-day lead time. Labeled tick marks in the bottom row highlight verification regions. For the 29 Jan storm: AL (Aleutians), YU (Yukon), and AN (Anchorage). For the 30 Jan storm: WA (Washington), OR (Oregon), CA (California), and CO (Colorado). For the 25 Feb storm: WC (West Coast), CA (California), CO (Colorado), and IA (Iowa).

all errors are calculated by first calculating the difference with respect to the T511 NR for each field of interest (e.g., TEE, SLP, precipitation, and geopotential height) and then averaging those values over the designated verification domain(s) at the designated verification date(s) for each experiment and the CTL. Then, this information is plotted as either an absolute value or a percent difference with respect to the CTL.

The Ideal_tquv experiment that assimilated all observation types (t , q , and uv) in the idealized domain shows the largest reduction in forecast error relative to the CTL, about 2%–6% greater than other experiments (see Fig. 5). Moreover, in terms of individual observation types, wind observations (Ideal_uv) were found to reduce the forecast error by the largest amount, followed by temperature (Ideal_t) and specific humidity (Ideal_q). This occurs for each storm, regardless of

whether the average is calculated with respect to forecast hour or verification date. Results are variable but significant for several time periods. For example, in all cases, the addition of perfect dropsonde observations over the idealized domain results in approximately a 10% improvement in forecast error within the first 3 days of the forecast. In Fig. 5a, the reduction in forecast error is greater than 20%, due to the idealized domain in Fig. 2, including all of Alaska. As a result, the initial forecast error is greatly reduced relative to CTL. For all storms, the impact of the data within the first 2–3 days of the forecast is generally statistically significant and positive.

In Figs. 5d–f, the average performance of all forecasts valid on certain dates is shown. Since most of the forecast error is likely due to storm activity, and the goal is to improve forecast of the storm, Figs. 5d–f are generated

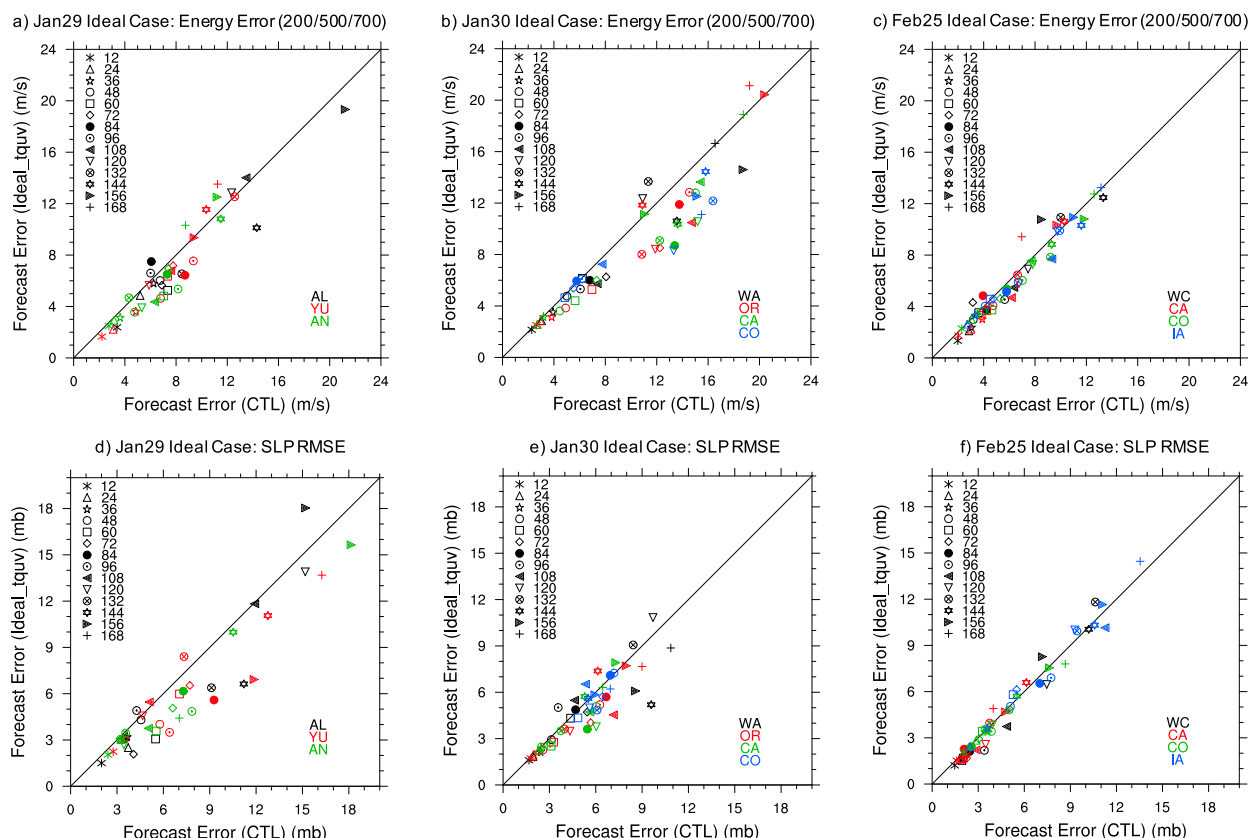


FIG. 6. Scatterplots of (top) TEE and (bottom) SLP RMSE for the CTL (x axis) and the Ideal_tquv experiment (y axis) for the three storms [(a),(d) 29 Jan; (b),(e) 30 Jan; and (c),(f) 25 Feb storms]. Errors are calculated for each indicated forecast lead time relative to each storm verification regions ($14^{\circ} \times 14^{\circ}$ lat–lon boxes, as specified in Fig. 2) and then averaged over the same regions. The colors represent the different verification regions (see Fig. 5 caption). The different symbols represent different forecast lead times relative to each verification region.

to highlight the change in forecast error for the storm itself. Though many of the results in these plots show a reduction in forecast error due to the assimilation of supplemental dropsonde observations, some show an increase in forecast error or no impact. This could be because DA is a statistical procedure that can have a lot of sampling noise due to the ensemble size.

For the 29 January storm, the impact of targeted observations on forecast error is positive and statistically significant from 0600 UTC 29 January to 1800 UTC 30 January for Ideal_tquv; from about 0000 UTC 1 February to 1800 UTC 2 February for Ideal_tquv and Ideal_uv (with Ideal_t and Ideal_q showing significant improvement during the last 12 h); and around 0000 UTC 4 February for all three experiments. At 1200 UTC 2 February, the storm is passing over Anchorage, Alaska, and therefore, this region could see an improvement of $\sim 10\%$ in forecast error, as shown in Fig. 5d. For the 30 January storm, results are significant for the following time periods: 0000 UTC to 1800 UTC 2 February for Ideal_tquv and Ideal_uv, and 1800

UTC 2 February to 1200 UTC 5 February for all three experiments, with Ideal_q shifting to insignificant for about a 24-h period in the middle of that time window. As a result, California (Colorado) could see up to a 13% (15%) reduction in forecast error. Around 1200 UTC 31 January, the variability is large for the 30 January storm for all experiments. At this time, there is an interaction between the ridge ahead of the storm and a cutoff low within the ridge, resulting in uncertainty in the forecasts. Figure 5f shows that results are significant for the 25 February storm at 1200 UTC 24 February for the Ideal_q experiment, at 1800 UTC 24 February for the Ideal_tquv experiment, and between 0000 UTC 27 February and 1800 UTC 28 February for the Ideal_tquv experiment (with Ideal_uv and Ideal_t also showing statistical significance during portions). An improvement in forecast error of $\sim 7\%$ at 0000 UTC 28 February for Ideal_tquv occurs for approximately a 24-h period when the storm is over the United States, moving eastward from California to Iowa. Additionally, the variability

is high from 0000 UTC 26 February to 0000 UTC 27 February (Fig. 5f) due to the storm momentarily becoming a closed low off the West Coast of the United States. An increase in forecast error due to a closed (25 February storm) or cutoff low (system downstream of 30 January storm) and their tendency to have no clear error propagation or forecast error growth is well known and documented in the forecast community (Gurka et al. 1995; Nieto et al. 2005; Caruso and Businger 2006; Nieto et al. 2008; Oakley and Redmond 2014). The results discussed here (see Fig. 5) do not change when using the 500-mb geopotential height RMSE instead of TEE as the metric for measuring changes in forecast error. For the 6-h accumulated precipitation, the differences in the errors are not statistically significant over the vast majority of forecast lead times and verification dates (not shown), such that the few exceptions that do exist occur at verification dates and lead times beyond the scope of this study. Precipitation will be discussed further in section 5.

For each storm, the forecast error is reduced during portions of its development due to the assimilation of perfect dropsonde observations upstream of its location. This result shows that targeted observations, when deployed over a large domain, can have a positive impact on forecast error for short-term weather forecasts.

c. Verification regions

Scatterplots of forecast error for each storm over each verification region (see Fig. 2) are presented in Fig. 6. This approach provides individual snapshots of forecast performance in and around each storm. In the following sections, only experiments with all dropsonde data types are evaluated, since the *tquv* experiment resulted in the largest overall reduction in forecast error and is the most realistic assimilation of dropsonde observations.

Figures 6a–c show that overall, TEE is reduced in the Ideal_*tquv* experiment relative to CTL for all storms during the first 3 to 4 days of the forecast, consistent with Fig. 5. Beyond day 4, results are generally neutral, with the exception of the 30 January storm. For the 30 January storm, forecast error decreases with additional dropsonde observations for almost all forecast lead times, up to approximately forecast hour 144. Overall, the largest reduction in forecast error for the 30 January storm is over the California verification region, with Colorado and Oregon also showing improvement, but to a lesser extent (consistent with Fig. 5e). For the 25 February storm, the largest (smallest) reduction in forecast error for Ideal_*tquv* relative to CTL occurs over California at shorter (longer) lead times (i.e., at times less than or greater than approximately forecast hour 84). The

opposite occurs for Colorado. This suggests that the reduction in forecast error for the Ideal_*tquv* experiment, as shown in Fig. 5f, may be the result of an early improvement in the forecast for California and a later improvement in the forecast for Colorado. Both the Pacific and Iowa verification regions see no significant change in forecast error (consistent with Fig. 5f). For the 29 January storm, Anchorage shows the most improvement, as also seen in Fig. 5d. Similar results are found in the 500-mb geopotential height RMSE (not shown) for all three storms. We also examined precipitation error for all three storms (not shown) and found no significant difference between the experiments and the CTL.

When SLP is used as a metric (Figs. 6d–f), results are similar to Figs. 6a–c. For the 29 January storm (Fig. 6d), the vast majority of forecast lead times for each verification region show a reduction in forecast error. This is in contrast to TEE, which is reduced for all regions for shorter lead times. For the 30 January storm, the decrease in the SLP RMSE (Fig. 6e) is reduced, compared to Fig. 6b, which results in the Colorado verification region shifting from some to no improvement in forecast error. Forecast error for the 25 February storm (Fig. 6f) is relatively unchanged between the two metrics. Overall, there is less variation in the data when using SLP, but the pattern and ranking of the verification regions is equivalent.

In general, the 30 January storm shows the most improvement in forecast error, followed by the 29 January storm, and then the 25 February storm, with neutral results. The overall neutral results for the 25 February storm could be the result of a multitude of things, such as 1) model physics errors preventing forecast improvements, 2) the resolution at which the GFS was run being too coarse to see a significant impact in smaller-scale processes, 3) the nonlinear nature of the atmosphere, and 4) the DA system not making the best use of the observations. To better understand why some storms seem to have greater improvement in forecast error than others, we investigate the sources of error for two of the storms, 25 February and 30 January, in relation to observation placement in the next section.

d. Sources of error

For this part of the analysis, we focused on areas outside the idealized domain shown in Fig. 2, where analysis errors would likely be larger (as shown in Figs. 3a,b and 4a,b) and could grow to negatively contribute to forecast error. For each storm, we isolated which meteorological structures entered and exited the idealized domain and at what time (see Table 2 for the 30 January storm and Table 3 for the 25 February storm). After these meteorological structures enter the idealized domain, the propagation of the data impact from the dropsonde data

TABLE 2. Meteorological features influencing the 30 Jan storm along with when each feature exits and enters the idealized sampling domain.

Feature	Enters sampling region	Exits sampling region
Cutoff low	Before 20 Jan	Begins: 1200 UTC 26 Jan 2006 Halfway: 2100 UTC 29 Jan 2006 Fully: 0000 UTC 31 Jan 2006
Ridge downstream of storm	Before 20 Jan	Begins: 1500 UTC 27 Jan 2006 Halfway: 1200 UTC 29 Jan 2006 Fully: 0000 UTC 31 Jan 2006
Main storm	Before 20 Jan	Begins: 0900 UTC 31 Jan 2006 Fully: 0600 UTC 1 Feb 2006
Ridge upstream of storm	Halfway: 1200 UTC 25 Jan 2006 Fully: 1800 UTC 26 Jan 2006	Around 1200 UTC 1 Feb 2006
Upper-level low over Canada	Never	Never

will be with the group velocity, which is faster than the phase velocity of the larger structures (Szunyogh et al. 2002). To understand why certain forecasts performed poorly, compared to the CTL, we tracked regions of high TEE back from a later forecast to the analysis time to find their origins and to isolate what meteorological structures most impacted forecast performance.

Figure 7 shows average forecast error based on TEE for the 30 January and 25 February storms and highlights the significant and consistent reduction in forecast error for the 30 January storm. For that storm, the forecast error is reduced for the Ideal_tquv experiment relative to CTL for forecasts initiated between 1200 UTC 27 January and 0000 UTC 2 February (see Figs. 7b–d). Prior to this date, the impact is neutral (see Fig. 7a), partially due to the cutoff low discussed in section 4b. Between 1200 UTC 27 January and 0000 UTC 2 February, all three synoptic structures identified in Table 2 are sampled by the simulated dropsondes. Further analysis of the horizontal propagation of TEE in and around the storm revealed that of the three structures, the ridge upstream of the storm had the most impact on the accurate forecast of the storm due to large errors associated with the developing ridge in the CTL (see Figs. 8a,b). This error starts to reduce in the

Ideal_tquv experiment (see Figs. 8c,d) when the ridge is completely in the idealized domain (beginning around 1800 UTC 26 January) and thus fully sampled by the simulated dropsondes. The importance of the ridge upstream of the storm could be tied to its later development into a Rex block pattern, which is a difficult weather pattern to predict due to its abrupt onset, resulting in an inherent sensitivity to blocking onset in NWP models (Pelly 2001; Lawson and Horel 2015). On the whole, most forecasts for the 30 January storm see a positive impact with the addition of simulated dropsonde observations, and this tendency is seen in the average error (Figs. 5, 6).

Forecast error plots for the 25 February storm reveal an overall neutral impact on forecast error for that storm when assimilating simulated dropsonde data into the NCEP GFS. The impact remains neutral until 1200 UTC 25 February, as shown in Fig. 7. Aside from the storm itself, three atmosphere structures were identified as having the potential to impact the development of the storm (see Table 3). When examining the propagation of TEE for individual forecasts, we found that the majority of the forecasts showed mixed results in forecast error over the western United States (see Figs. 7e,f) due to a short wave downstream of the storm. In this case, the development of the short wave is influenced by

TABLE 3. Meteorological features influencing the 25 Feb storm along with when each feature exits and enters the idealized sampling domain.

Feature	Enters sampling region	Exits sampling region
Ridge downstream of storm	Before 20 Feb	0000 UTC 25 Feb 2006
Main storm	Halfway: 0000 UTC 21 Feb 2006 Fully: 0000 UTC 22 Feb 2006	0000 UTC 27 Feb 2006
Ridge upstream of storm	1200 UTC 23 Feb 2006	0000 UTC 28 Feb 2006
Short wave	0600 UTC 25 Feb 2006	1800 UTC 28 Feb 2006

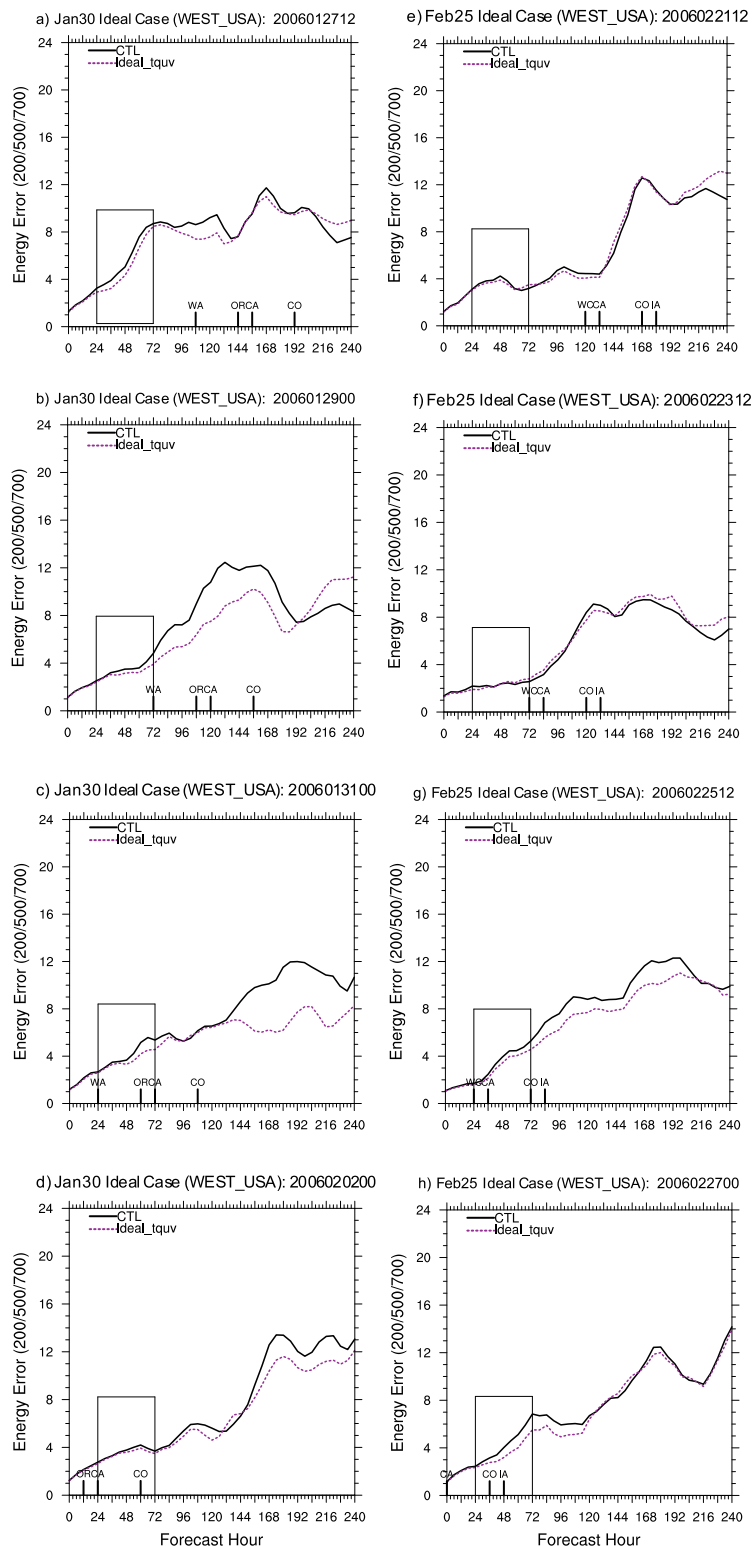


FIG. 7. Plots of average forecast error, measured as TEE, for the CTL (black solid line) and Ideal_tquv (purple dashed line) over the western United States for the (left) 30 Jan and (right) 25 Feb storms. Only select forecast initialization times, specified in the title of each plot, are shown to highlight certain patterns of interest as discussed in the text.

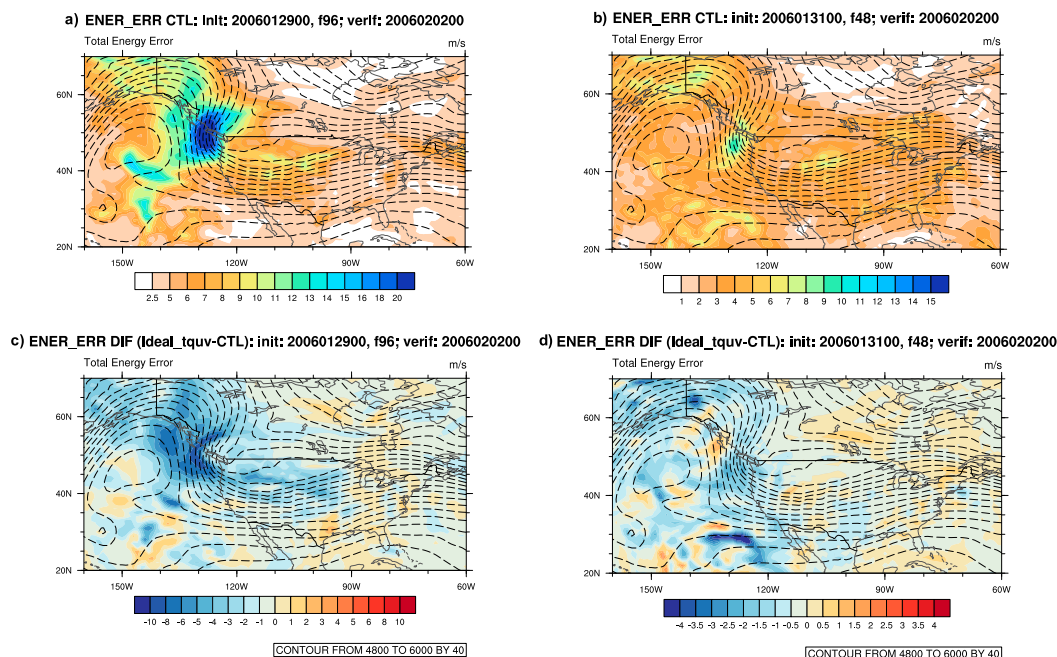


FIG. 8. Spatial distribution of (top) CTL forecast TEE and (bottom) the difference between CTL and Ideal_tquv TEE for two forecast initialization times [(a),(c) 0000 UTC 29 Jan and (b),(d) 0000 UTC 31 Jan 2006]. Both initialization times have a verification date of 0000 UTC 2 Feb 2006. Line contours are 500-mb geopotential heights showing the location of the storm. Note difference in color bars due to larger errors in longer-range forecasts.

atmospheric structures not sampled (e.g., upper-level low over eastern Canada) and undersampled (e.g., ridge downstream of storm and short wave), resulting in neutral impact. Additionally, forecasts initiated before 1200 UTC 25 February assimilated observations before the storm became a closed low, which contributed to the mixed results seen before this date. Around a forecast initialization time of 1200 UTC 25 February, the Ideal_tquv experiment begins to show a consistent reduction in forecast error relative to the CTL, as shown in Fig. 7g. This reduction is found to be due to sampling the embedded short wave upstream of the storm and the storm when it temporarily becomes a closed low (Figs. 9a,c). Additionally, results show that the short wave has the greater impact on forecast error, and, as a consequence, sampling that short wave results in a larger reduction in forecast error (Figs. 9b,d).

5. Results: Sensitivity and Flight experiments

In the previous experiments, a large idealized domain was sampled, and simulated dropsonde observations were produced. This is not realistic for targeted observations since flight campaigns can only sample a portion of the atmosphere over a short period of time. To evaluate the impact of a more realistic domain, we looked at two smaller

sampling domains: a sensitivity and a flight domain. For the Sensitivity and Flight experiments, instead of cycling for multiple days, as in the Idealized experiments, cycling is performed for five cycles or approximately a 24-h period due to the maximum flight duration of the GH. The flight occurs 2–3 days prior to the storm reaching the verification region (0000 UTC 30–31 January for the 29 January storm, 1200 UTC 30–31 January for the 30 January storm, and 0000 UTC 24–25 February for the 25 February storm).

a. ETS maps

Maps of normalized ETS values (Fig. 10) highlight the sensitive regions that should be targeted to reduce the forecast errors over a verification region with a specific lead time. Each map is an average of five cycles in order to capture the movement of the signal over a 2–3-day range and to average out noise. The ETS maps vary for each storm and each verification region, and thus, one verification region was chosen for each storm: Anchorage (AN) for the 29 January storm, Oregon (OR) for the 30 January storm, and California (CA) for the 25 February storm. Verification regions were selected using the following four factors for guidance while not requiring that all were satisfied: 1) meteorological impact, 2) the potential for improvement in the CTL, 3) how much the forecast improved in the Idealized

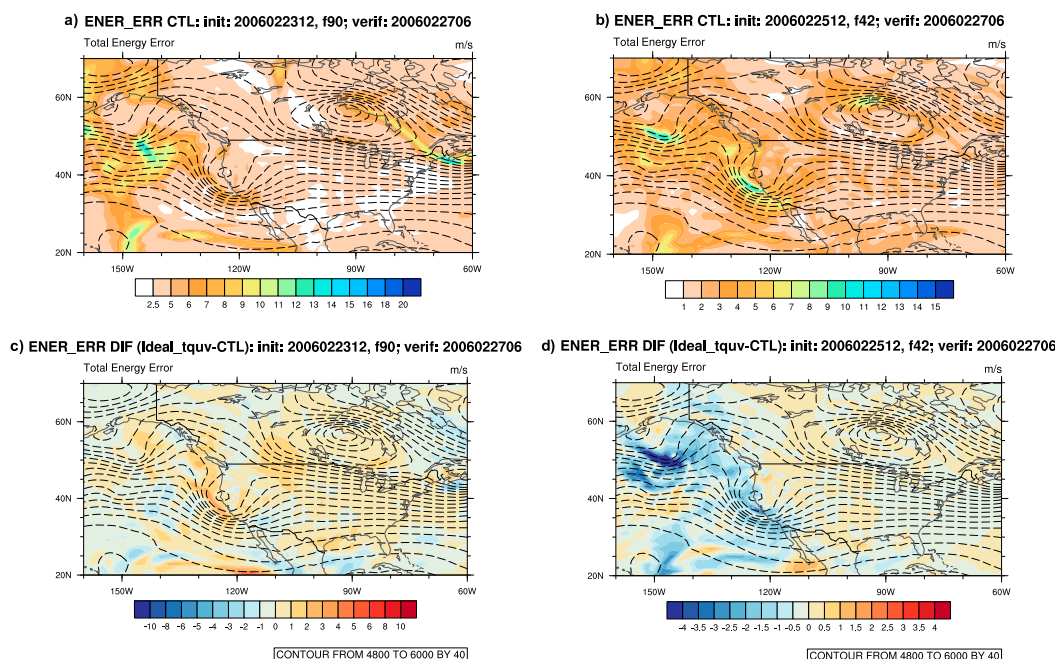


FIG. 9. As in Fig. 8, but for the 25 Feb storm with forecast initialization times of (left) 1200 UTC 23 Feb and (right) 1200 UTC 25 Feb 2006 and a verification date of 0600 UTC 27 Feb 2006.

experiments, and 4) the spatial pattern of the sensitive regions shown in the ETS maps. The sensitivity domain is defined using Fig. 10, such that simulated observations for the Sensitivity experiments are generated where the ETS is greater than or equal to 0.5 (locations enclosed by the white contour lines in Fig. 10). This method results in a total of 94, 765, and 349 dropsondes being released for the 29 January, 30 January, and 25 February storms, respectively.

In Fig. 10, the red dots represent the location of simulated dropsonde observations along the flight path generated for each storm. The number of dropsondes deployed and the flight time for the Flight experiments are in the titles of each plot. The flight path domain is defined by creating simulated flight paths over the sensitivity domain based upon aircraft capabilities, as described in section 3. Specific flight paths are generated using an automated algorithm to place simulated dropsondes over regions with high ETS values (see appendix C).

b. Analysis and forecast performance

Before evaluating any change in forecast error, the impact of the targeted dropsonde observations on the analysis error was assessed. Figure 11 shows the change in the 500-mb zonal wind RMSE for the analysis after the assimilation of additional dropsonde observations over the sensitivity (Figs. 11a–c) and flight (Figs. 11d–f) domains for each of the three storms. In these plots, the analysis error is reduced over the selected regions

(Fig. 10) for both sets of experiments. Plots generated on other pressure levels (200 and 700 mb) and for other variables (temperature, SLP, and geopotential height) show similar results (not shown).

Percent change in TEE for the three sampling domains (idealized, sensitivity, and flight) for forecasts initialized at 2–3-day lead times for each storm are shown in Fig. 12. As in Fig. 5, the solid lines in this figure represent the percent mean difference, and the dashed lines represent the confidence interval. When comparing the verification regions of interest for each storm, the 29 January storm shows the largest reduction in forecast error for the Idealized experiment (TEE of -22% , SLP of -1.2 mb) and the least for the Sensitivity and Flight experiments, with the forecast error for the Sensitivity experiments not significant. The 30 January storm has the second-largest reduction in forecast error for the Idealized experiment (TEE of -16% ; SLP of -0.4 mb) and the largest reduction in forecast error for the Sensitivity and Flight experiments (TEE of -3% and -2% ; SLP of -0.15 and -0.5 mb, respectively). Results for this storm are statistically significant at the 95% confidence interval for all three experiments. Finally, the 25 February storm has the smallest reduction in forecast error for all experiments (TEE of -7% for Idealized, -2% for Sensitivity, and -1% for Flight; SLP of $+0.2$ mb for Idealized and no change for the other experiments), with none of the results being statistically significant. Between the West Coast and California

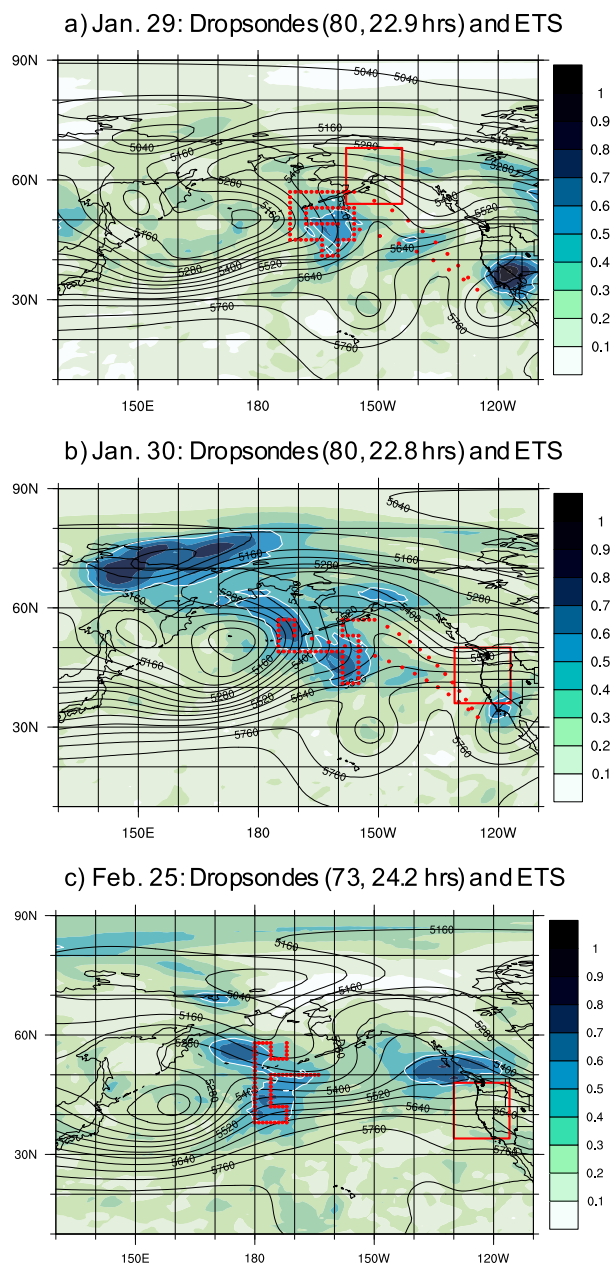


FIG. 10. Maps of sensitivity indicating regions from where error growth is expected to influence a verification region the most. Regions are identified using the ETS method and then normalized and averaged over a 24-h period (or five cycles) 2–3 days prior to the verification time. The verification region, shown with the red box for each storm, is (a) Anchorage for the 29 Jan storm, (b) Oregon for the 30 Jan storm, and (c) California for the 25 Feb storm. For the Sensitivity experiments, simulated observations are generated where the ETS value is 0.5 (white contour line) or greater. The red dots represent the location of simulated dropsonde observations along the flight path generated for each storm. The title shows the number of dropsondes deployed and the flight time for each Flight experiment.

verification regions for the 25 February storm, there is an increase in the forecast error variability (dashed line) and a decrease in the mean forecast error (solid line) for the Idealized experiment. As before, this is likely due to the storm temporarily becoming a closed low before returning to the mean flow. During this same time period, the Idealized experiment shows less improvement, compared to the other two experiments. This is due to the closed low deepening much more in the Idealized experiment than the NR for one of the five forecasts, which results in large errors from one forecast decreasing the overall forecast improvement. Results were also generated for 500-mb geopotential heights and 6-h accumulated precipitation (not shown). Plots of the differences in the geopotential height RMSE were found to have a pattern and significance similar to that of TEE, whereas precipitation was not statistically significant.

The Idealized experiment in Fig. 12 has the largest reduction in forecast error. The Sensitivity and Flight experiments generally result in small, but statistically significant, reductions in forecast error. Overall, the assimilation of targeted observations is found to result in a statistically significant positive impact over the following regions (see Figs. 12a–c): 1) Anchorage for the 29 January storm for the Idealized experiment, 2) Oregon for the 30 January storm for all three experiments, and 3) California for the 25 February storm for the Idealized and Sensitivity experiments. Plots of SLP bias in Fig. 12 result in the same general conclusions.

Scatterplots of forecast error for each storm over each verification region (see Fig. 2) are presented in Fig. 13. As before, the Idealized experiments perform the best for all storms, showing the largest reduction in forecast error. Additionally, SLP RMSE and TEE results shown in Figs. 13a,b,d,e for both January storms are consistent with Fig. 12. For the 25 February storm, the TEE for the Sensitivity and Flight experiments increases relative to the CTL (Fig. 13c), which does not match the small reduction in forecast error shown in Fig. 12c. However, the SLP RMSE for the Idealized experiment decreases in Fig. 12f, which does not match the increase in forecast error seen in Fig. 12f. This could be because there can be significant variability in individual results, which is evident in Fig. 6, where many data points fall on the zero line, but there are many others that show an improvement in forecast error. Precipitation bias is also evaluated and found to be highly variable from storm to storm (see Figs. 13g–i). The 30 January storm improved the most, with the reduction in forecast error being concentrated at the longer lead times (66 and 72 h). The 25 February storm also improved, but at the shorter forecast lead times (48 and 54 h). These changes in

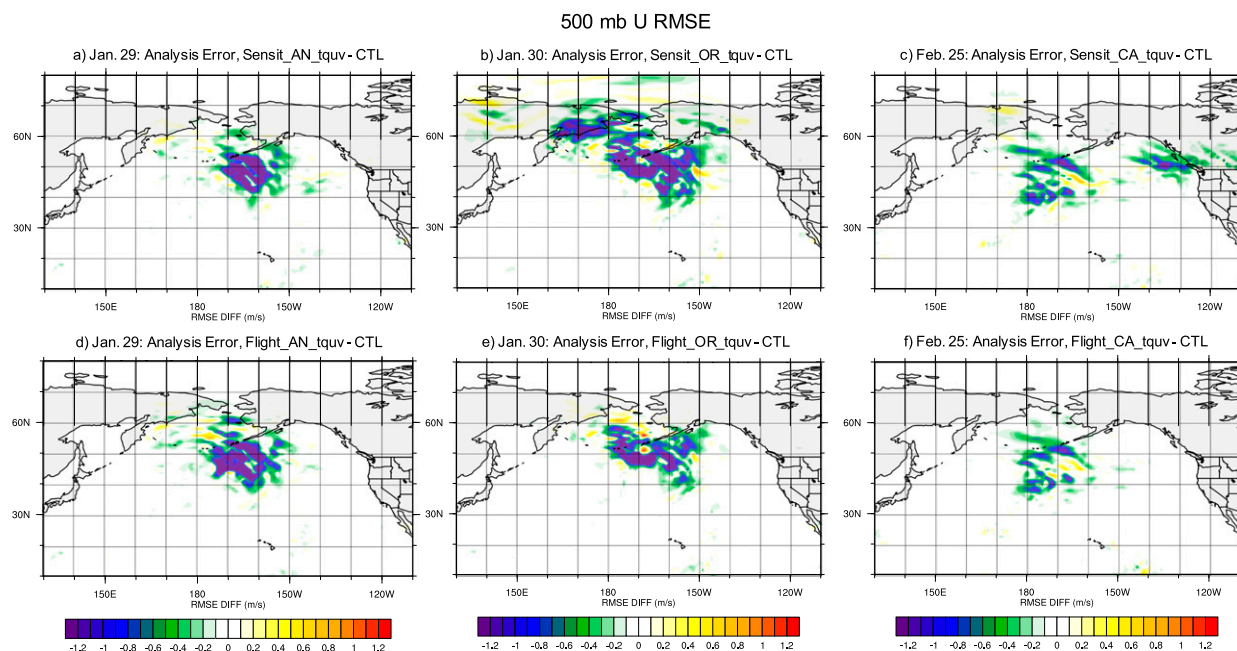


FIG. 11. Change in the 500-mb zonal wind analysis RMSE after dropsondes are assimilated for the (top) Sensitivity and (bottom) Flight experiments and for (a),(d) 29 Jan; (b),(e) 30 Jan; and (c),(f) 25 Feb storms.

precipitation are small and found to not be statistically significant (not shown).

Overall, the 29 January storm does not seem to benefit from targeted observations, as shown by the Sensitivity and Flight experiments, whereas the 25 February storm does benefit. The 30 January storm has the largest reduction in forecast error and improves with all experiments. These results highlight the positive, but case-dependent, impact of targeted observations on forecast error in a real-time campaign scenario.

6. Discussion

A review by Majumdar (2016) of targeted observations discusses multiple real-world campaigns, some of which were focused on extratropical cyclones. Even though that review contains only OSE experiments, it is still informative to compare those results to the OSSE results in this manuscript since they tend to overestimate the forecast error reduction. The North Pacific Experiment (NORPEX-98) campaign sought to improve 1–3-day winter storm forecasts over North America and sampled the atmosphere with about 700 dropwindsondes over the campaign, resulting in the assimilation of approximately 26 soundings per forecast (Langland et al. 1999a). Results showed an average reduction in forecast error for the targeted cases of about 10% for 500-mb geopotential heights. This real-world result is greater than the approximate 5% (3%) reduction seen in the Sensitivity (Flight) experiments

using simulated observations for 500-mb geopotential heights (not shown). Another campaign, the WSR program, sought to improve 1–5-day winter storm forecasts over North America (e.g., Szunyogh et al. 2000; Toth et al. 2000; Szunyogh et al. 2002; Holland et al. 2004). For WSR99, about 500 dropwindsondes were deployed over 19 flights, resulting in about 26 dropwindsondes per flight assimilated into a low-resolution (T62) version of the NCEP operational 3D-Var assimilation system (Szunyogh et al. 2000). Szunyogh et al. (2000) found that the largest surface pressure RMSEs were reduced by 10%–20% (or 0.5 mb). Similar results were found for the 2000 WSR program, as presented in Szunyogh et al. (2002). When looking at 2-day lead times for our work, the SLP RMSE reduced by up to 0.2 mb for both the Sensitivity and Flight experiments (see Figs. 12d–f). Hamill et al. (2013) also examined the impact of targeted observations using observations from the WSR program, but in a 4DVar system, and found neutral impacts on average. However, only about eight dropsondes (776 over program) per flight were assimilated, which is much fewer than the approximately 70 dropsondes assimilated per cycle in this OSSE study.

One reason for the differing results between these past studies and the Sensitivity and Flight experiments could be due to the ability of the Hybrid 3DENVar T382L64 system to more effectively propagate information from data-rich to data-poor areas and thus is not as vulnerable to data gaps (Kelly et al. 2007). As a

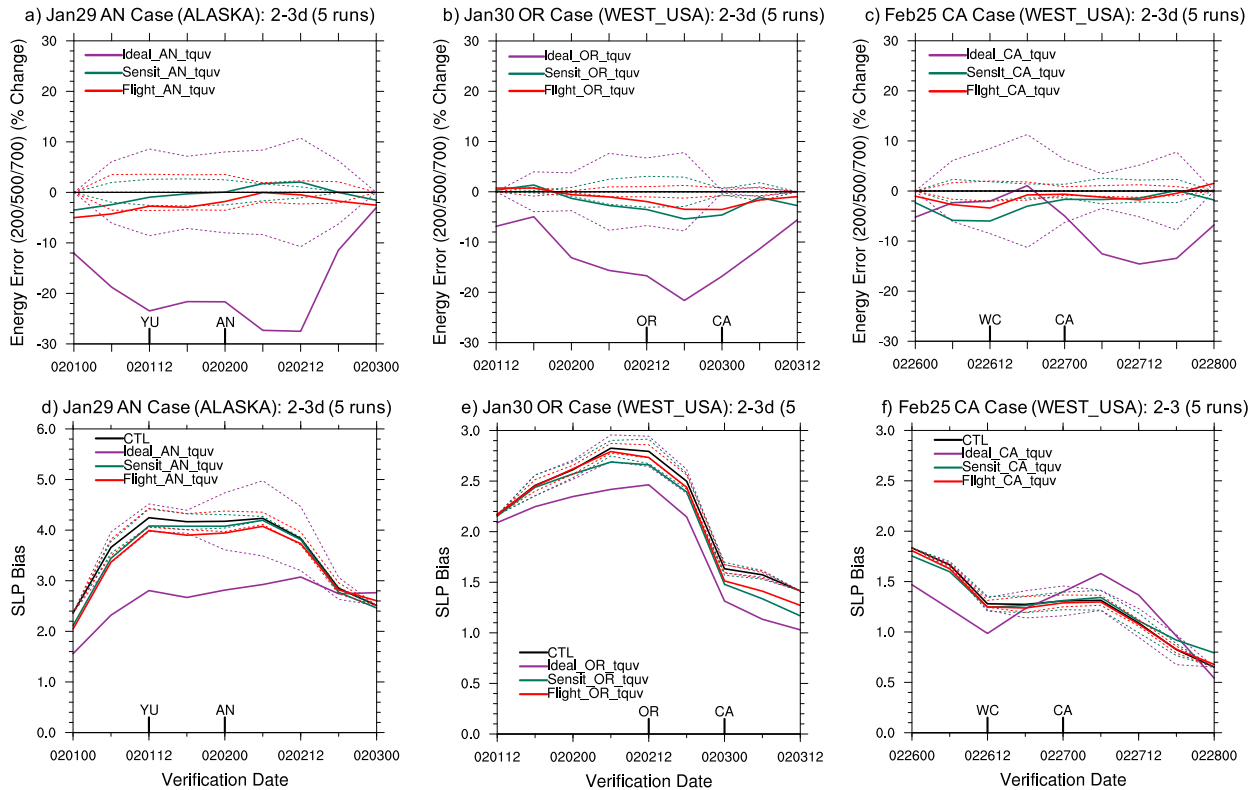


FIG. 12. (top) TEE and (bottom) SLP bias error for each set of experiments for the (a),(d) 29 Jan, (b),(e) 30 Jan, and (c),(f) 25 Feb storms. Notice that the range of the y axis for the (e) 30 Jan and (f) 25 Feb plots of SLP bias is different than that for (d) 29 Jan. Error values are generated by calculating errors for all forecasts with a 2–3-day lead time relative to one verification region ($14^{\circ} \times 14^{\circ}$ lat–lon boxes outlined in black in Fig. 2) and then averaging those errors for each storm over the domain of interest (either WEST_USA or ALASKA). The percent difference between each experiment and the CTL is shown with the solid lines and the 95% confidence interval (obtained using the paired t test) with the dashed lines, such that Ideal_(region)_tqv is in purple, Sensit_(region)_tqv in green, and Flight_(region)_tqv in red (see Fig. 5 caption for more details). For all plots, the standard deviation goes to zero at the edges of the plot because $n = 1$ at these points. Moving toward the center of the plot, the number of 2–3-day forecasts used in the calculation increases, such that n is maximized and equals 5. Labeled tick marks highlight verification regions. For the 29 Jan storm: YU (Yukon) and AN (Anchorage). For the 30 Jan storm: OR (Oregon) and CA (California). For the 25 Feb storm: WC (West Coast) and CA (California).

result, supplemental observations may not have the same impact on forecast skill as seen in an inferior assimilation system (Kelly et al. 2007). For example, work by Buizza et al. (2007) found a 4% reduction in forecast error verified over North America using ECMWF 4D-Var system when studying the impact of Pacific Ocean observations. This is much lower than previous studies that have used a 3D-Var nonhybrid system (e.g., Langland et al. 1999a; Szunyogh et al. 2000, 2002) but closer to the results shown in the Sensitivity and Flight experiments presented in this study. However, as presented in section 1, there are some differing viewpoints as to the impact of targeted observations in different DA systems. Other reasons that results from this work differ from others, in addition to that already mentioned, could be case selection, verification domains, targeting technique,

meteorological situation, season, year, and DA system used (e.g., Kelly et al. 2007; Buizza et al. 2007; Cardinali et al. 2007).

As expected, this study found that the Ideal_tqv experiment performs the best and that the dropsonde observations have the greatest (least) positive impact on forecast error for the 30 January (25 February) storm. Section 4d explored forecast error growth in both of these storms and found that both had multiple structures that contributed to this growth, with the strongest feature associated with a ridge upstream of the storm (i.e., Rex block for the 30 January storm and an embedded short wave for the 25 February storm). These structures are also important for the Sensitivity and Flight experiments, since the ridge upstream of the 30 January storm and the associated propagation of the observation signal (Szunyogh et al. 2000) are captured using the

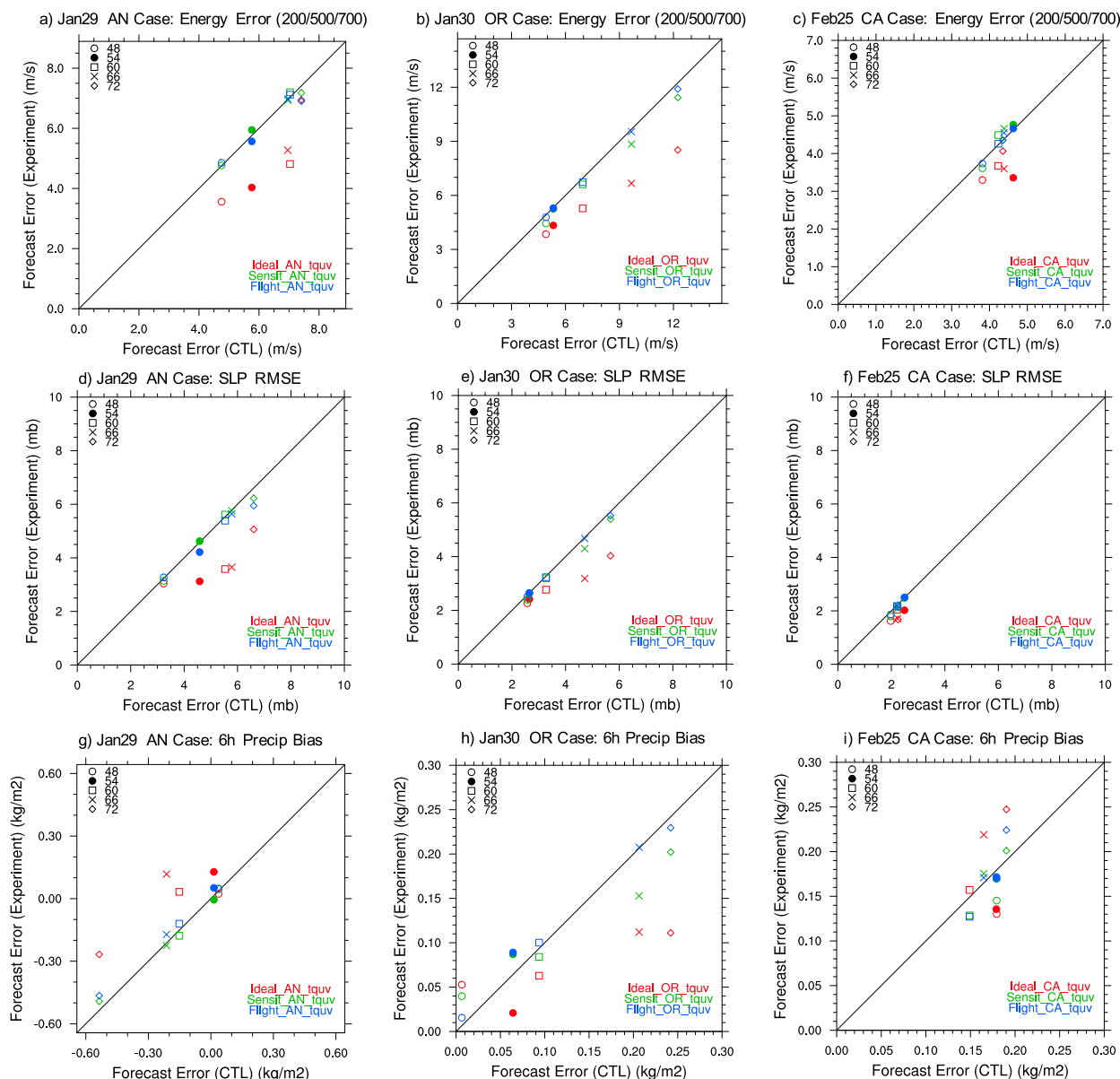


FIG. 13. Scatterplots of (top) TEE, (middle) SLP RMSE, and (bottom) 6-h accumulated precipitation bias for the CTL (x axis) and each experiment (y axis) for the 29 Jan, 30 Jan, and 25 Feb storms. Values are generated by calculating the respective error for forecasts with a 2–3-day lead time relative to one verification region and then averaging those errors for each storm over the same region ($14^{\circ} \times 14^{\circ}$ lat-lon boxes outlined in black in Fig. 2). For each storm, there are symbols representing the forecast lead time (48–72 h) for each set of experiments: Ideal (red), Sensit (green), and Flight (blue). The verification region for each storm is shown in the plot titles and in the legend with AN (Anchorage), OR (Oregon), and CA (California).

ETS technique. However, the short wave and the storm becoming a closed low is not captured for the 25 February storm since all the forecasts are initiated before these two important structures formed. Instead, the ETS technique captures the deepening upper-level low from which the 25 February storm subsequently develops.

It is important to think about the potential for forecast improvement for each storm, the selection of cases with

clear perturbation propagation and amplification, the size of the sampling region, and the intensity and spatial extent of the sensitive regions found by the ETS technique. In the Idealized experiments, the analysis error is larger (of equal intensity but more spatially extensive) for the 30 January storm (see Figs. 3a,b), compared to the 25 February storm (see Figs. 4a,b). These analysis errors produce poorer forecasts seen in the CTL. For the Sensitivity experiments (see Fig. 10), the sensitivity

domain for the 30 January storm has the largest spatial extent with large relative ETS values (the same is true when examining absolute ETS values, not shown) when compared to the other two storms. Therefore, the 30 January storm, which has large analysis errors and sensitivity, has higher potential for improvement. The 29 January storm has the smallest spatial extent and magnitude for relative ETS values, but the absolute ETS values are between that of the other two storms with the weakest error propagation signal (not shown). The weaker signal could partially explain why the Sensitivity experiment produces a 3% reduction in TEE for the 30 January storm, and the 29 January storm (see Fig. 12) produces neutral results even though the absolute ETS values are stronger than the 25 February storm. The 25 February storm appears mixed in terms of improvement in forecast error (see Fig. 12). This could also be partly due to the characteristics of the sensitivity domain since the domain has a small spatial extent with low relative ETS values, but higher absolute ETS values with a clearer error propagation signal (not shown). The Flight experiments generally show an even smaller reduction in forecast error, compared to the Sensitivity experiments, due to their smaller domain.

7. Conclusions

Extratropical cyclones can greatly impact the United States and cause significant loss of life and property due to intense winds and heavy precipitation. Even though NWP has improved, forecast busts still occur in cases of strong North Pacific cyclones (McMurdie and Mass 2004; McMurdie and Casola 2009; Rodwell et al. 2013) due to large analysis errors over data-sparse regions, such as the Pacific Ocean (Langland et al. 1999b, 2002; McMurdie and Casola 2009; Colle and Charles 2011). In this study, a targeted observation approach is used within the OSSE framework to improve forecast skill of winter storms impacting the western United States. Three U.S. winter storms are identified in the ECMWF T511 NR between January and February 2006. The ETS method is used to find regions sensitive to forecast error growth (Zhang et al. 2016). Three different experiments with different combinations of dropsonde observations were designed and compared to a CTL without dropsonde observations (see Table 1): 1) Idealized, 2) Sensitivity, and 3) Flight experiments providing dense sampling of (i) a very large domain, (ii) areas sensitive to error growth, and (iii) a flight track over the sensitive areas found in (ii).

The primary results from this study are the following:

- 1) Initializing the GFS with simulated perfect observations over a large idealized domain of the Pacific

Ocean improves forecasts over the CONUS for all three winter storms studied, at times reducing energy error by up to 15%. Even though these experiments use perfect simulated observations (no errors added prior to being assimilated) and underrepresent the real world, the impact is large, indicating the potential value of supplemental observations even with realistic simulated observations (perfect + errors added prior to assimilation into GSI).

- 2) Sampling all three observation types (temperature, wind, and specific humidity) provides the most benefit. Individually, temperature and wind are found to provide more benefit than specific humidity. This could be due to using the dry energy-norm or to specific humidity being less important for the scales and processes in this study.
- 3) The ETS technique is effective at isolating regions of high sensitivity that could reduce the 2–3-day forecast error over specific verification regions, if sampled.
- 4) At the 2–3-day lead time, sampling the sensitivity or flight path domain with simulated perfect observations provides a reduction in TEE of up to 5% (a smaller reduction in forecast error than sampling the idealized domain, as expected). Even though this positive impact would tend to be reduced with nonperfect observations, this study still suggests that targeted observations can provide an improvement in forecasts.
- 5) Reduction in forecast error of an average of 10% for the Idealized experiments is comparable to previous studies that used lower-resolution 3D-Var nonhybrid DA system. A reduction of up to 5% for the Sensitivity and Flight experiments is less than that found in previous studies but comparable to a more recent study that found a 4% reduction using a modern 4D-Var model with real observations.
- 6) For all experiments, the 30 January storm shows a larger reduction in forecast error than the 25 February storm, compared to the CTL. This is potentially due to the undersampling of important atmospheric structures critical for storm development and the spatial characteristics of the analysis error available in the CTL.

Acknowledgments. This work is funded by NOAA Award NA14OAR4830170 as part of the SHOUT project and conducted under the auspices of NOAA's Quantitative Observing System Assessment Program (QOSAP). Drs. Tanya Peevey and Jason English were supported by funding from NOAA Award NA17OAR4320101 and Dr. Hongli Wang was supported by funding from NOAA Award Number NA14OAR4320125. Thanks to colleagues at NOAA/OAR/ESRL/GSD and NOAA/AOML/HRD for their support and stimulating conversations. We would specifically like to thank Dr. Zoltan

TABLE A1. Satellite observations assimilated into the NCEP GFS system for this OSSE study.

Satellite	Microwave	Infrared	Radio occultation
<i>Aqua</i>	AMSU-A	AIRS	
<i>MetOp-A</i>	AMSU-A, MHS	HIRS4, IASI	GRAS
<i>Suomi NPP</i>	ATMS	CrIS	
<i>NOAA-15</i>	AMSU-A		
<i>NOAA-18</i>	AMSU-A, MHS		
<i>NOAA-19</i>	AMSU-A, MHS	HIRS4	
<i>Meteosat-9</i>		SEVIRI	
<i>GOES-13</i>		SNDR D1–D4	
<i>DMSP F16</i>	SSMIS		
GRACE			JPL Blackjack
COSMIC			JPL Blackjack
<i>TerraSAR-X</i>			JPL Blackjack
<i>C/NOFS</i>			CORISS

Toth and Dr. Yuanfu Xie for their advice on the implementation of the ETS method, Dr. Sean Casey for providing the simulated conventional and satellite data, and Kate Friedman at the NCEP for her assis-

tance in setting up the NCEP GFS Model on the Zeus and Theia supercomputers. Special thanks to Dr. Bob Atlas for his recommendations on evaluation metrics of forecast skill.

APPENDIX A

Simulated Observations

Both simulated satellite and conventional observations were used in this OSSE study. A list of simulated observations assimilated into the GSI system is shown in [Tables A1](#) and [A2](#). This list was motivated by the need to be as close to NCEP operations as possible so that the results could help to inform decision-makers and support the public good. As a result, the list of satellites and conventional observations assimilated was set by the NCEP operations usage files.

APPENDIX B

Comparison of Dry Total-Energy Norms

In this section, [Fig. B1](#) is presented and compared to [Fig. 5](#). Forecast error in [Fig. B1](#) is generated using the dry total-energy norm from [Hamill et al. \(2013\)](#) instead of the one presented in this manuscript. A comparison of [Fig. 5](#) and [Fig. B1](#) highlights the equivalency of the two

TABLE A2. Conventional observations assimilated into the NCEP GFS system for this OSSE study. In this table, T_v and T_s are virtual and sensible temperature, q is specific humidity, u and v are zonal and meridional wind, and P_{stn} is station pressure.

PREPBUFR message type	Description	Report type	Observations
ADPUPA	Rawinsonde	120 (220)	$T_v, q, P_{\text{stn}}(u, v)$
	Flight-level reconnaissance and profile dropsonde	132 (232)	$T_v, q(u, v)$
	Pilot balloon (pibal)	221	u, v
AIRCAR	Meteorological Data Collection and Reporting System (MDCRS) Aircraft Communications, Addressing, and Reporting System (ACARS) aircraft	133 (233)	$T_s, q(u, v)$
AIRCFT	Aircraft report (AIREP) and pilot report (PIREP) aircraft	130 (230)	$T_s(u, v)$
	Aircraft Meteorological Data Relay (AMDAR) aircraft	131 (231)	$T_s(u, v)$
SFCSHP	Surface marine reports	180 (280)	$T_v, q, P_{\text{stn}}(u, v)$
	Splash level drop over ocean	182 (282)	$T_v, q, P_{\text{stn}}(u, v)$
ADPSFC	Surface land (SYNOP, METAR) reports	181	T_v, q, P_{stn}
PROFLR	NOAA Profiler Network	223	u, v
SATWND	Satellite wind reports	242–243, 245–46, 250, 252–54, 257–59	u, v
WDSATR	WindSat scatterometer data	289	u, v
ASCATW	Advanced Scatterometer (ASCAT) winds over ocean	290	u, v
VADWND	NEXRAD VAD wind data	224	u, v

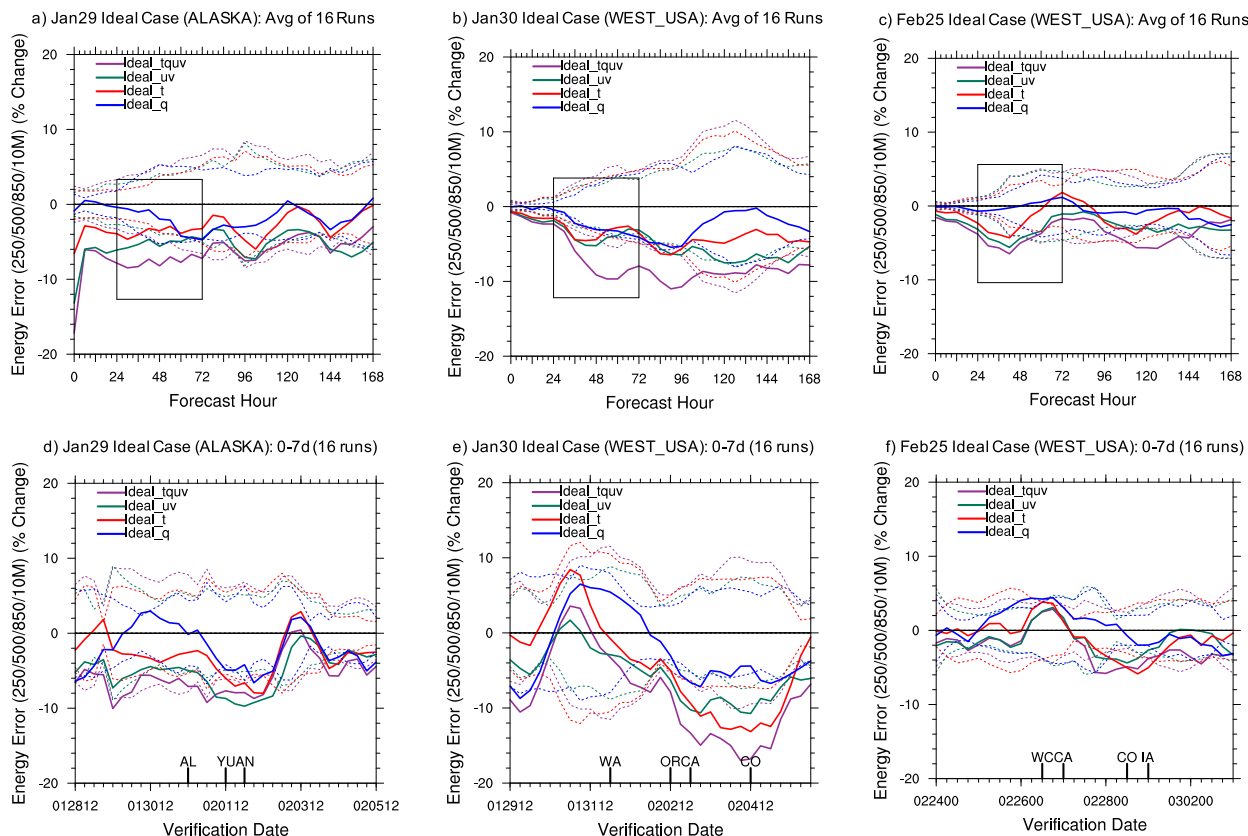


FIG. B1. As in Fig. 5, but generated with the dry total-energy norm used in Hamill et al. (2013).

formulations, indicating that the surface fields are minor contributors to the dry total-energy norm for the storms analyzed in this study. Overall, the pattern and relative changes between each experiment for each storm is about the same in these figures. More specifically, the reduction in forecast error for the experiments relative to the CTL is strongest at shorter forecast lead times, with the Idealized experiment having the largest reduction. Additionally, verification regions identified as having statistically significant results for each experiment are the same between both figures (see bottom rows). Exceptions to this are found around 0000 UTC 30 January for the 30 January storm and 0000 UTC 25 February for the 25 February storm, when the statistical significance for some of the experiments changes from significant to insignificant. These verification times are beyond the verification regions/dates of interest presented in this study in section 5. For those verification regions, the differences between using the dry total-energy norm presented here and that from Hamill et al. (2013) may be significant, and if it was not beyond the scope of this study, would have been investigated further.

APPENDIX C

Automated Flight Track Algorithm

All UAS flight tracks took into account the ETS sensitivity maps. First off, we had to account for the GH flight requirements: 1) only fly over the ocean due to Federal Aviation Administration (FAA) regulations, 2) maximum flight duration of ~ 24 h, 3) up to 80 dropsondes per GH flight, 4) travel as far west as the central North Pacific (180°) per FAA GH constraints, and 5) distance from departure location at the NASA's Armstrong Flight Research Center at Edwards Air Force base in California.

After accounting for the GH flight constraints, the track software generates a GH flight path that considers the 2–3-day average ETS sensitivity pattern. From the sensitivity, the track software creates an initial waypoint of the largest ETS sensitivity in the specified GH domain that is 7 h from Edwards Air Force Base. Although this time is arbitrary, it was chosen to maximize flight time over the sensitive region for releasing dropsonde data. Future waypoints are chosen based on the gradient of the ETS sensitivity in the north, south, west, and east directions from the initial waypoint.

From this perspective, the flight path traverses the region of the greatest sensitivity in order to maximize the sampling of regions prone to large error growth. The flight path ends when the maximum allowable time limit is reached. When that happens, a final waypoint is computed at Edwards Air Force Base. From the waypoints produced, dropsonde locations are created at a density of $\pm 1^\circ$ of latitude–longitude in the primary sampling domain. The algorithm does not take into account the atmospheric flow or the nature run wind field. One of the future considerations is to account for the atmospheric wind as well as important meteorological features tied to the sensitive areas. The algorithm, however, provides an objective method to sample the region of highest ETS sensitivity.

REFERENCES

- Aberson, S., 2008: Large forecast degradations due to synoptic surveillance during the 2004 and 2005 hurricane seasons. *Mon. Wea. Rev.*, **136**, 3138–3150, <https://doi.org/10.1175/2007MWR2192.1>.
- , 2010: 10 years of hurricane synoptic surveillance (1997–2006). *Mon. Wea. Rev.*, **138**, 1536–1549, <https://doi.org/10.1175/2009MWR3090.1>.
- , S. Majumdar, C. Reynolds, and B. Etherton, 2011: An observing system experiment for tropical cyclone targeting techniques using the Global Forecast System. *Mon. Wea. Rev.*, **139**, 895–907, <https://doi.org/10.1175/2010MWR3397.1>.
- Andersson, E., and M. Masutani, 2010: Collaboration on Observing System Simulation Experiments (Joint OSSE). *ECMWF Newsletter*, No. 123, ECMWF, Reading, United Kingdom, 14–16.
- Atlas, R., and Coauthors, 2015: Observing system simulation experiments (OSSEs) to evaluate the potential impact of an optical autocovariance wind lidar (OAWL) on numerical weather prediction. *J. Atmos. Oceanic Technol.*, **32**, 1593–1613, <https://doi.org/10.1175/JTECH-D-15-0038.1>.
- Bishop, C., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765, [https://doi.org/10.1175/1520-0469\(1999\)056<1748:ETAAO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<1748:ETAAO>2.0.CO;2).
- Boukabara, S.-A., and Coauthors, 2016: Community Global observing system simulation experiment (OSSE) Package (CGOP): Description and usage. *J. Atmos. Oceanic Technol.*, **33**, 1759–1777, <https://doi.org/10.1175/JTECH-D-16-0012.1>.
- Buizza, R., C. Cardinali, G. Kelly, and J.-N. Thépaut, 2007: The value of observations. II: The value of observations located in singular-vector-based target areas. *Quart. J. Roy. Meteor. Soc.*, **133**, 1817–1832, <https://doi.org/10.1002/qj.149>.
- Cardinali, C., R. Buizza, G. Kelly, M. Shapiro, and J.-N. Thépaut, 2007: The value of observations. III: Influence of weather regimes on targeting. *Quart. J. Roy. Meteor. Soc.*, **133**, 1833–1842, <https://doi.org/10.1002/qj.148>.
- Caruso, S., and S. Businger, 2006: Subtropical cyclogenesis over the central North Pacific. *Wea. Forecasting*, **21**, 193–205, <https://doi.org/10.1175/WAF914.1>.
- Charles, M., and B. Colle, 2009: Verification of extratropical cyclones within the NCEP operational models. Part I: Analysis errors and short-term NAM and GFS forecasts. *Wea. Forecasting*, **24**, 1173–1190, <https://doi.org/10.1175/2009WAF2222169.1>.
- Colle, B., and M. Charles, 2011: Spatial distribution and evolution of extratropical cyclone errors over North America and its adjacent oceans in the NCEP Global Forecast System model. *Wea. Forecasting*, **26**, 129–149, <https://doi.org/10.1175/2010WAF2222422.1>.
- Cucurull, L., R. Li, and T. R. Peevey, 2017: Assessment of radio occultation observations from the COSMIC-2 mission with a simplified observing system simulation experiment configuration. *Mon. Wea. Rev.*, **145**, 3581–3597, <https://doi.org/10.1175/MWR-D-16-0475.1>.
- English, J., A. Kren, and T. Peevey, 2018: Improving winter storm forecasts with observing system simulation experiments (OSSEs): Part 2, Evaluating a satellite gap with idealized and targeted dropsondes. *Earth Space Sci.*, <https://doi.org/10.1002/2017EA000350>.
- Errico, R., R. Yang, N. Privé, K.-T. Tai, R. Todling, M. Sienkiewicz, and J. Guo, 2013: Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1162–1178, <https://doi.org/10.1002/qj.2027>.
- Gurka, J., E. Auciello, A. Gigi, J. Waldstreicher, K. Keeter, S. Businger, and L. Lee, 1995: Winter weather forecasting throughout the eastern United States. Part II: An operational perspective of cyclogenesis. *Wea. Forecasting*, **10**, 21–41, [https://doi.org/10.1175/1520-0434\(1995\)010<0021:WWFTTE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0021:WWFTTE>2.0.CO;2).
- Hamill, T., F. Yang, C. Cardinali, and S. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Mon. Wea. Rev.*, **141**, 2058–2065, <https://doi.org/10.1175/MWR-D-12-00309.1>.
- Hoffman, R., and R. Atlas, 2016: Future observing system simulation experiments. *Bull. Amer. Meteor. Soc.*, **97**, 1601–1616, <https://doi.org/10.1175/BAMS-D-15-00200.1>.
- Holland, L., Z. Toth, J. Moskaitis, S. Majumdar, C. Bishop, and R. Smith, 2004: Results from the Winter Storm Reconnaissance program 2002–2003. *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, Seattle, WA, Amer. Meteor. Soc., 7.8, https://ams.confex.com/ams/84Annual/techprogram/paper_72177.htm.
- Hu, M., H. Shao, D. Stark, K. Newman, C. Zhou, G. Ge, and X. Zhang, 2017: Grid-point Statistical Interpolation (GSI) user's guide version 3.6. Developmental Testbed Center Rep., 149 pp., <http://www.dtcenter.org/com-GSI/users/docs/index.php>.
- Huang, L., and Z. Meng, 2014: Quality of the target area for metrics with different nonlinearities in a mesoscale convective system. *Mon. Wea. Rev.*, **142**, 2379–2397, <https://doi.org/10.1175/MWR-D-13-00244.1>.
- Irvine, E., S. Gray, J. Methven, I. Renfrew, K. Bovis, and R. Swinbank, 2009: The impact of targeted observations made during the Greenland Flow Distortion Experiment. *Quart. J. Roy. Meteor. Soc.*, **135**, 2012–2029, <https://doi.org/10.1002/qj.499>.
- , —, —, and —, 2011: Forecast impact of targeted observations: Sensitivity to observation error and proximity to steep orography. *Mon. Wea. Rev.*, **139**, 69–78, <https://doi.org/10.1175/2010MWR3459.1>.
- Kelly, G., J.-N. Thépaut, R. Buizza, and C. Cardinali, 2007: The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Quart. J. Roy. Meteor. Soc.*, **133**, 1803–1815, <https://doi.org/10.1002/qj.150>.
- Kleist, D., 2012: An evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Ph.D. Thesis, University of Maryland, Dept. of Atmospheric and Oceanic Science, 163 pp.
- , and M. C. Morgan, 2005: Application of adjoint-derived forecast sensitivities to the 24–25 January 2000 U.S. East

- Coast snowstorm. *Mon. Wea. Rev.*, **133**, 3148–3175, <https://doi.org/10.1175/MWR3023.1>.
- , and K. Ide, 2015a: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, <https://doi.org/10.1175/MWR-D-13-00351.1>.
- , and —, 2015b: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4D-EnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, <https://doi.org/10.1175/MWR-D-13-00350.1>.
- Langland, R., R. Gelaro, G. Rohaly, and M. Shapiro, 1999a: Targeted observations in FASTEX: Adjoint-based targeting procedures and data impact experiments in IOP17 and IOP18. *Quart. J. Roy. Meteor. Soc.*, **125**, 3241–3270, <https://doi.org/10.1002/qj.49712556107>.
- , and Coauthors, 1999b: The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Amer. Meteor. Soc.*, **80**, 1363–1384, [https://doi.org/10.1175/1520-0477\(1999\)080<1363:TNPEX>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<1363:TNPEX>2.0.CO;2).
- , M. Shapiro, and R. Gelaro, 2002: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 East Coast snowstorm. *Mon. Wea. Rev.*, **130**, 957–974, [https://doi.org/10.1175/1520-0493\(2002\)130<0957:ICSAEG>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0957:ICSAEG>2.0.CO;2).
- Lawson, J., and J. Horel, 2015: Ensemble forecast uncertainty of the 1 December 2011 Wasatch windstorm. *Wea. Forecasting*, **30**, 1749–1761, <https://doi.org/10.1175/WAF-D-15-0034.1>.
- Ma, Z., L. Riishøjgaard, M. Masutani, J. Woollen, and G. Emmitt, 2015: Impact of different satellite wind lidar telescope configurations on NCEP GFS forecast skill in observing system simulation experiments. *J. Atmos. Oceanic Technol.*, **32**, 478–495, <https://doi.org/10.1175/JTECH-D-14-00057.1>.
- Majumdar, S., 2016: A review of targeted observations. *Bull. Amer. Meteor. Soc.*, **97**, 2287–2303, <https://doi.org/10.1175/BAMS-D-14-00259.1>.
- Masutani, M., and Coauthors, 2007: Progress in joint OSSEs. *18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 12B.5, <https://ams.confex.com/ams/pdfpapers/124080.pdf>.
- , and Coauthors, 2010: Observing system simulation experiments at the National Centers for Environmental Prediction. *J. Geophys. Res.*, **115**, D07101, <https://doi.org/10.1029/2009JD012528>.
- McCarty, W., R. Errico, and R. Gelaro, 2012: Cloud coverage in the joint OSSE nature run. *Mon. Wea. Rev.*, **140**, 1863–1871, <https://doi.org/10.1175/MWR-D-11-00131.1>.
- McClung, T., 2014: NWS Technical Implementation Notice 14-46, <http://www.nws.noaa.gov/om/notification/tin14-46gfs.htm>.
- McMurdie, L., and C. Mass, 2004: Major numerical forecast failures over the Northeast Pacific. *Wea. Forecasting*, **19**, 338–356, [https://doi.org/10.1175/1520-0434\(2004\)019<0338:MNFOT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0338:MNFOT>2.0.CO;2).
- , and J. Casola, 2009: Weather regimes and forecast errors in the Pacific Northwest. *Wea. Forecasting*, **24**, 829–842, <https://doi.org/10.1175/2008WAF2222172.1>.
- Morss, R., K. Emanuel, and C. Snyder, 2001: Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.*, **58**, 210–232, [https://doi.org/10.1175/1520-0469\(2001\)058<0210:IAOSFI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<0210:IAOSFI>2.0.CO;2).
- Newton, C., 1954: Analysis and data problems in relation to numerical prediction. *Bull. Amer. Meteor. Soc.*, **35**, 287–294.
- Nieto, R., and Coauthors, 2005: Climatological features of cutoff low systems in the Northern Hemisphere. *J. Climate*, **18**, 3085–3103, <https://doi.org/10.1175/JCLI3386.1>.
- , M. Sprenger, H. Wernli, R. Trigo, and L. Gimeno, 2008: Identification and climatology of cut-off lows near the tropopause. *Ann. N. Y. Acad. Sci.*, **1146**, 256–290, <https://doi.org/10.1196/annals.1446.016>.
- Oakley, N., and K. Redmond, 2014: A climatology of 500-hPa closed lows in the northeastern Pacific Ocean, 1948–2011. *J. Appl. Meteor. Climatol.*, **53**, 1578–1592, <https://doi.org/10.1175/JAMC-D-13-0223.1>.
- Palmer, T., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653, [https://doi.org/10.1175/1520-0469\(1998\)055<0633:SVMAAO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2).
- Pelly, J., 2001: The predictability of atmospheric blocking. Ph.D. dissertation, University of Reading, 207 pp.
- Privé, N., R. Errico, and K.-S. Tai, 2014a: The impact of increased frequency of rawinsonde observations on forecast skill investigated with an observing system simulation experiment. *Mon. Wea. Rev.*, **142**, 1823–1834, <https://doi.org/10.1175/MWR-D-13-00237.1>.
- , Y. Xie, S. Koch, R. Atlas, S. Majumdar, and R. Hoffman, 2014b: An observing system simulation experiment for the unmanned aircraft system data impact on tropical cyclone track forecasts. *Mon. Wea. Rev.*, **142**, 4357–4363, <https://doi.org/10.1175/MWR-D-14-00197.1>.
- Reale, O., L. Terry, M. Masutani, E. Andersson, L. Riishøjgaard, and J. Jusem, 2007: Preliminary evaluation of the European Centre for Medium-Range Weather Forecasts (ECMWF) Nature Run over the tropical Atlantic and African monsoon region. *Geophys. Res. Lett.*, **34**, L22810, <https://doi.org/10.1029/2007GL031640>.
- Riishøjgaard, L., Z. Ma, M. Masutani, J. Woollen, G. Emmitt, S. Wood, and S. Greco, 2012: Observation system simulation experiments for a global wind observing sounder. *Geophys. Res. Lett.*, **39**, L17805, <https://doi.org/10.1029/2012GL051814>.
- Rodwell, M., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, <https://doi.org/10.1175/BAMS-D-12-00099.1>.
- Szunyogh, I., Z. Toth, K. Emanuel, C. Bishop, C. Snyder, R. Morss, J. Woollen, and T. Marchok, 1999: Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.*, **125**, 3189–3217, <https://doi.org/10.1002/qj.49712556105>.
- , —, R. Morss, S. Majumdar, B. Etherton, and C. Bishop, 2000: The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance program. *Mon. Wea. Rev.*, **128**, 3520–3537, [https://doi.org/10.1175/1520-0493\(2000\)128<3520:TEOTDO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3520:TEOTDO>2.0.CO;2).
- , —, A. Zimin, S. Majumdar, and A. Persson, 2002: Propagation of the effect of targeted observations: The 2000 Winter Storm Reconnaissance program. *Mon. Wea. Rev.*, **130**, 1144–1165, [https://doi.org/10.1175/1520-0493\(2002\)130<1144:POTEOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1144:POTEOT>2.0.CO;2).
- Toth, Z., I. Szunyogh, S. Majumdar, R. Morss, B. Etherton, C. Bishop, and S. Lord, 2000: Targeted observations at NCEP: Toward an operational implementation. *Fourth Symp. on Integrated Observing Systems*, Long Beach, CA, Amer. Meteor. Soc., 186–193.
- , and Coauthors, 2002: Adaptive observations at NCEP: Past, present, and future. *Symp. on Observations, Data Assimilation, and Probabilistic Prediction*, Orlando, FL, Amer. Meteor. Soc., 185–190.
- Wang, X., 2010: Incorporating ensemble covariance in the grid-point statistical interpolation variational minimization: A

- mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, <https://doi.org/10.1175/2010MWR3245.1>.
- , D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, <https://doi.org/10.1175/MWR-D-12-00141.1>.
- Wedam, G., L. McMurdie, and C. Mass, 2009: Comparisons of model forecast skill over the East and West Coasts of the United States. *Wea. Forecasting*, **24**, 843–854, <https://doi.org/10.1175/2008WAF2222161.1>.
- Whitaker, J., and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- , and —, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- , —, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, <https://doi.org/10.1175/2007MWR2018.1>.
- Zhang, Y., Y. Xie, H. Wang, D. Chen, and Z. Toth, 2016: Ensemble transform sensitivity method for adaptive observations. *Adv. Atmos. Sci.*, **33**, 10–20, <https://doi.org/10.1007/s00376-015-5031-9>.
- Zhu, T., F. Weng, M. Masutani, and J. S. Woollen, 2012: Synthetic radiance simulation and evaluation for a Joint Observing System Simulation Experiment. *J. Geophys. Res.*, **117**, D23111, <https://doi.org/10.1029/2012JD017697>.